

A time for scientific diplomacy

Both India and Pakistan have much to gain from closer collaboration between their scientific communities. Such collaboration must not be allowed to remain a casualty of tensions between the two.

The decisions by India and Pakistan to detonate nuclear devices will have few — if any — direct consequences for the relations between ordinary citizens of both countries. They won't, for example, diminish the passion Pakistanis have for Indian cinema, nor the fondness that Indians hold for Pakistani television soap operas. India won't stop Muslims from Pakistan flocking to the shrines of revered Sufi saints. And Pakistan won't stop Sikhs from India crossing the border to visit their second-holiest temple.

But the tests will almost certainly hamper communication between one group of Indians and Pakistanis who would like the chance to talk more: the scientists of both countries. The irony for those who gave their countries nuclear weapons status is that they face increased isolation, not only from the West, but also from each other. Science is perhaps one of the few vehicles that could help raise both the quality of life and levels of trust between these two quarrelsome neighbours. But at present, official, bilateral scientific cooperation does not exist.

Scientists from both countries are not totally isolated from one another. They meet at international venues, such as the Abdus Salam Centre for Theoretical Physics in Trieste, Italy, and at United Nations environment conventions, where both countries form part of the Group of 77 developing states. Pakistan's scientists can, in theory, travel to New Delhi to visit the International Centre for Genetic Engineering and Biotechnology. Indian scientists, similarly, can travel to Islamabad to visit the headquarters of the intergovernmental Commission on Science for Sustainable Development for the South. Some maintain good personal relations with cross-border colleagues.

But in practice, contact is rare. Scientific collaboration may have

helped thaw the Cold War in the West, but it has become one of the casualties of the continued tense relations between India and Pakistan. Politicians from both sides view science as a key element of each country's defence and security, and consider scientific cooperation — no matter how innocuous — as close to giving away state secrets. If pressed, they tend to take the view that enhanced scientific collaboration will follow progress on outstanding political issues, not vice versa.

This is unfortunate, as scientific collaboration can — as the West has shown — help to ease political tensions. It is also wrong, as such collaboration could bring urgent, practical benefits to both countries. Indeed, science could be harnessed to help India and Pakistan tackle a range of common problems. These include diseases such as malaria and tuberculosis; agricultural issues such as developing salt tolerance in crops; and environmental issues such as air pollution, as India and Pakistan have similar types of road transport.

There is, in fact, much evidence that scientists from both countries would like to collaborate. But they need both permission and support from their politicians, which has so far been lacking. This support in turn needs courage, diplomacy, an element of 'lateral thinking', and an enlightened view of India–Pakistan relations.

Such thinking is not unknown and once came from an unlikely source. In 1987, Pakistan's military president General Zia ul Haq tried to defuse rising tensions by travelling across the border to make an unscheduled appearance at an India–Pakistan cricket match. The incident was coined 'cricket diplomacy'. And it worked. Pakistan's present prime minister, Nawaz Sharif, a Zia protégé, could take a cue from his mentor, and engage in a bit of 'scientific diplomacy'. What does he have to lose? □

When moderation turns to greed

The NIH must be ready to take firm action to protect access to the tools that are essential to a researcher's work.

Few issues are as frustrating to the contemporary researcher, particularly in the life sciences, as the time and effort spent negotiating the terms on which he or she can gain access to the research tools required to do an efficient job. In a less complex — perhaps slightly mythical — era, biological samples were exchanged freely between laboratories, equipment manufacturers were content merely if their products carried out specified tasks effectively, and experimental results were disseminated widely and rapidly.

In today's world, as a report compiled by a working group set up by the National Institutes of Health (NIH) makes clear (see page 505), the demands of licensing agreements and the enforced need to respect the intellectual property of others is reaching critical dimensions. Samples can be jealously guarded, manufacturers can exploit the ownership of unique technologies, the terms of collaboration can be stringent, and publication can be delayed on demand. The working party's conclusion is stark: "We believe that current trends pose a serious threat to the best interests of the biomedical research and development community as a whole."

The problem is when common sense and moderation give way to

greed. It is one thing, for example, to expect a reasonable return from a discovery that will undoubtedly bring profit to others; it is another to exploit the virtual monopoly position which the law, in the hands of an aggressive and imaginative patent attorney, can potentially allow one researcher to exercise over another. The fault does not always lie in the explicitly profit-motivated private sector. The NIH working group says it found those in industry who had been taken aback by the price of cooperation being demanded by some university researchers.

Given the diversity of situations in which tensions can arise, and the widely varying practices between different parts of the scientific community, there are no easy solutions — certainly none that could be enshrined in compact legislation. The NIH, also aware of the limited range of its responsibilities, is right to talk more in terms of building a consensus around broad guidelines. But it must not shirk from making unpopular judgements, for example by exercising its 'march-in' rights, where the long-term interests of biomedical research conflict with the short-term avarice of individuals. Science is one domain where belief in Adam Smith's 'invisible hand' can create as many problems as it solves. □



New Russian funding cuts spark protests from researchers

[MOSCOW] The Russian cabinet is proposing to cut science funding this year by 26.5 per cent compared with the figures approved by the State Duma — the lower house of the Russian parliament — in January.

The planned reduction — from Rubl 11.2 billion (\$1.8 billion) to Rubl 8.2 billion — which forms part of a broad programme of public economies, has already led to strong protests from Russian scientists. They picketed the finance ministry in Moscow last week, and plan to block roads into the nation's capital next week.

The protest is being organized by trade unions representing researchers, who have jointly formed the Russian co-ordinating committee of the scientific collectives (RCC).

RCC representatives have met with first deputy minister Valery Kostyuk, who is acting minister of science and technologies during the absence, due to illness, of the recently appointed minister Vladimir Bulgak. Kostyuk promised that the cabinet would start negotiations with RCC after 10 June.

This is the date on which Bulgak is scheduled to lead a governmental commission made up of representatives of his ministry, as well as the ministries of finances, economics and state property. But in a statement, RCC says its meeting with Kostyuk failed to reach agreement on a single issue of importance.

"This further reduction of state financing suggests that there will be a mass liquidation of scientific collectives, with a lot of people engaged in science losing their jobs, and eliminating any prospect of scientists' receiving the higher wages promised to them earlier," says RCC chairman Valery Sobolev.

His deputy, Aleksey Zharov, is even more outspoken: "This sequestering of the budget means the final and irrevocable destruction of Russian science."

But the cabinet's decision, which led to the RCC protest, is only the tip of the iceberg. Earlier last month, the government issued a decree indicating that the budget for science will experience even heavier cuts over the next few years.

In 2001, for example, under the government's austerity plans, spending on science will be 0.23 per cent of gross national product, compared with 0.33 in 1997, falling from 2.1 to 1.8 per cent of public expenditure.

The Ministry of Finances has already written to all research institutes and laboratories emphasizing that strict economies are to be imposed. From next year, state spending on scientific research will be cut by Rubl 1.5 billion each year until 2001.

Among those likely to suffer are the Russian

Foundation for Fundamental Research and the Russian humanitarian scientific foundations, both of which will see their current budgets cut by half.

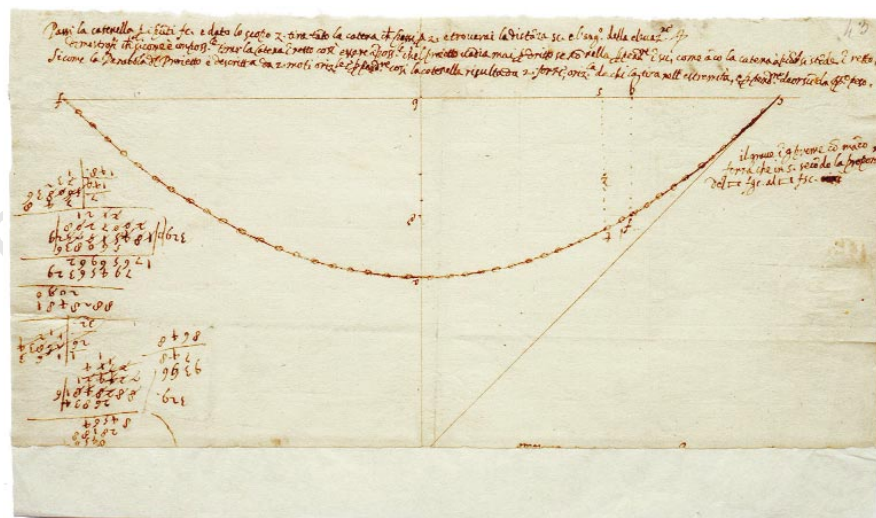
Furthermore, research institutes will be required to give priority to paying their electricity bills and other communal costs, leaving staff salaries to be paid out of what money — if any — remains.

The budget reduction affects all of the research programme except space research for which, it was reported, the cabinet has managed to find an extra Rubl 1.5 billion.

The government's moves have come in the wake of various proposals it has received that there should be a substantial pruning of research institutes and laboratories, to ensure that only the most competitive survive.

So far, however, efforts to move in this direction have been strongly resisted by the Russian Academy of Sciences. "It is easy to destroy something, but very difficult to build," says Andrey Gonchar, RAS vice-president. "What will be the long-term value of closing institutes and laboratories, even if they are currently ineffective?" **Carl Levitin**

Galileo's manuscripts go on the Internet



Web page: many of Galileo's manuscripts are now on the Internet, including this one showing an experimental protocol for checking his theory of projectile motion.

[MUNICH] Galileo Galilei's personal manuscripts, in which he wrote down the ideas, calculations and drawings that led to his theory of mechanics, can now be viewed on the Internet. It is believed to be the first historical scientific document of such significance to be made available in this way.

The 300-odd Internet folios are high-resolution photographs of Galileo's loose-leaf notes — in effect, his laboratory notebooks — from the late 1580s, when he was a young professor at the University of Pisa and still a follower of Aristotle, to 1638, when he published his *Discorsi* on mechanics. This publication marked a turning point in the development of scientific reasoning and opened the door to modern mechanics.

The Internet publication is a combined project by the Italian National Library of Florence, where most of Galileo's original texts are held, the Institute and Museum for

the History of Science in Florence, and the Max Planck Institute for the History of Science in Berlin. All of the Latin and Italian text in the manuscripts has been transcribed and is provided in Italian and English.

The text of the *Discorsi* is also on the web site and there are hypertext links between the notes and relevant pages in the *Discorsi*. The website will be modified in response to input from Galileo scholars.

"The Internet is changing the relationship between scholars, libraries and publishers," says Jürgen Renn, director of the Berlin Max Planck Institute. Web access means scholars no longer need to travel to the National Library in Florence to access Galileo's work. Renn says plans to include the collected works of Albert Einstein were abandoned because of copyright problems.

The texts are available on

<http://www.mpiwg-berlin.mpg.de> or

<http://www.imss.fi.it>

Alison Abbott

Report urges US to spend more money on R, but not D

[WASHINGTON] The US government should support more basic research at universities, rationalize its own laboratories, and avoid supporting technology development, according to a study by business leaders and academics.

The study, 'America's Basic Research', was conducted by the Committee for Economic Development (CED), an industry-backed think-tank. It argues that government support for basic research is critically important to US industry, but that technology development ought to be paid for by industrial corporations.

The report applauds the performance of the research universities but singles out the laboratories of the Department of Energy for criticism. They "have not acted forcefully to eliminate work in areas no longer relevant to their missions, nor to expose themselves to merit-based peer-review processes," the report claims.

But William Beeman, director of economic studies at CED, admits that the report's authors did not consult the energy department about these criticisms. Department officials argue that considerable progress has been made since similar criticisms were voiced in a celebrated 1995 report from a panel chaired by Robert Galvin, former chairman of the Motorola electronics corporation. "We were essentially endorsing the views of others who have looked at the laboratories," Beeman admits.

He adds that the study's main finding is that government support for science, and not technology, matters to the panel members. The panel started by looking broadly at science and technology, he says, but ended up concentrating on basic research because that was what mattered to the industrialists on the panel.

"They are not concerned about support for technology development," Beeman says. "They reject the argument that it is needed to ensure international competitiveness — and most academics reject that too."

These conclusions were immediately contested by Senator Jay Rockefeller (Democrat, West Virginia), an advocate of technology programmes, whom CED asked to speak at its report's launch. "I remain convinced that these programmes fill a very clear void that industry would not fill on its own," he said.

The CED report condemns the widespread congressional practice of 'earmarking' research money for projects at specified locations. It also calls for guidelines to ensure that relationships with industry do not damage the "primary basic research mission" of universities.

Colin Macilwain

Framework programme to get new advisory system

[MUNICH] The European Commission in Brussels is revamping the structure of its research advisory system in order to improve the management of its Fifth five-year Framework Research programme (FP5), which is due to begin next year.

The new structure includes the setting up of external advisory groups to oversee each of the 17 or so 'key actions' — the detailed research areas into which FP5 is divided. A call for nominations for membership of these groups should be published by the commission this week.

There will also be a 'two-chamber' body, one representing industry and the other the academic community, to advise on European Union research policy in general.

The commission sees this as a restructuring of its two existing advisory bodies, the European Science and Technology Assembly (ESTA), a 61-strong group of scientists and industrialists set up in 1994 to advise on the commission's research policy, and the 25-member Industrial R&D Advisory Committee (IRDAC), set up in 1984.

The commission says that a new system is necessary because, unlike the current framework programme of research, FP5 is closely focused on key actions and socio-economic objectives. According to a commission spokesman, the close relationship in FP5 between the discoverers of knowledge — primarily researchers — and those exploiting that knowledge — primarily industry and consumers — needs to be reflected in its advisory structure.

The research ministers of EU member states, who met informally in April to discuss the commission's research management practices (see *Nature* 392, 849; 1998), are keen that the advisory groups should be in place by the summer to provide input into the key actions' working programmes before 'calls for proposals' are put out at the end of the year.

Most of the key actions will have their own external advisory group made up of 15 to 20 members from the industrial and research communities, as well as relevant stakeholders. However, the four key actions associated with FP5's information technology programme will share one advisory group because of their integrated nature.

The groups will follow progress within their key actions to ensure that objectives are being met. They will advise on appropriate criteria for evaluating the actions, and, when felt necessary, on how their aims should be re-orientated.

The chairmen of each group will be members of one of the two proposed chambers.

Other members will be nominated by the commission, which says there will be "a balanced representation of the various research disciplines and economic sectors as well as an appropriate geographical and gender distribution".

The commission wants the chambers in place by the end of the year. They would share a secretariat in Brussels, and would be expected to work jointly or separately on topics that would be either set by the commission or proposed by the chambers themselves. Details of how it would work in practice are yet to be worked out.

The new system is said to have the support of the EU research ministers, although their formal approval is not required. There is also support from IRDAC, which last year suggested that the commission should combine its advisory forces given FP5's highly target-orientated structure.

But within ESTA, which is generally thought to have less influence than IRDAC on EU research policy, feelings are mixed.

Alexandre Quintanilha, a member of ESTA's 11-strong bureau, which discussed the issue with commission officials last week, says the bureau responded to the proposed two-chamber system with "guarded optimism".

Quintanilha, who is director of the Institute for Molecular and Cellular Biology in Oporto, Portugal, says the system could work well on topics such as technology transfer. But he says that ESTA members are worried that the views of the academic community may be overwhelmed by those of industry.

Frank Gannon, however, director of the European Molecular Biology Organization, welcomes the proposed changes, arguing that ESTA was always in danger of being marginalized because of FP5's strong orientation towards product development.

Jan Borgmann, ESTA's first chairman, says he is "surprised that the changes to the commission's advisory structures are happening so quickly and are so radical", given that ESTA has only been established for four years, and its new assembly has only been in place for six months. He is also concerned that the commission's desire to secure a geographical balance in the two chambers may weaken their influence.

ESTA's current assembly was appointed for a two-year period last autumn, and has since been deliberating on how to interact with the commission over issues relating to FP5. Other activities under consideration have been suspended until decisions are made about its future.

Alison Abbott

UK's Dounreay reprocessing plant to shut

[LONDON] The British government has announced that its nuclear-waste reprocessing plant at Dounreay, Scotland, once the heart of the UK fast-reactor programme, is to close within the decade. The plant will cease to take further reprocessing orders with immediate effect.

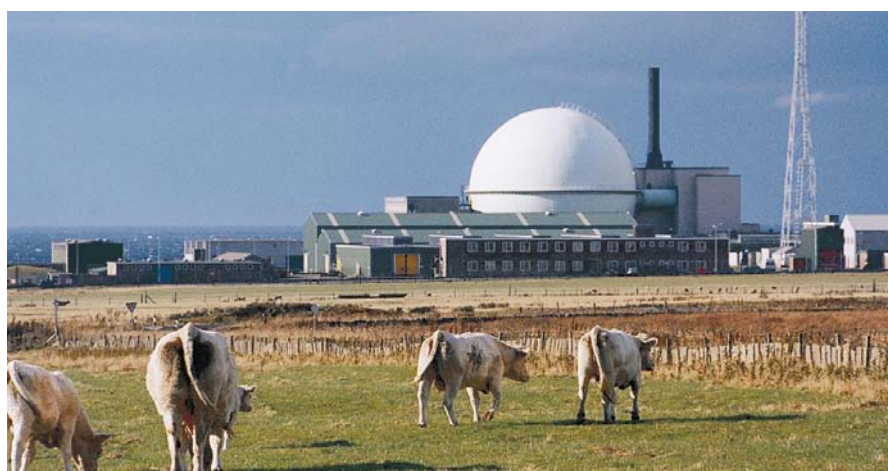
A commitment to existing contracts — including reprocessing a consignment of highly enriched uranium from Georgia (see *Nature* 392, 850; 1998) — will keep Dounreay open until 2006. After then, the government wants to turn the site into a model for decommissioning nuclear plants.

However, it is anticipated that redundancies are unlikely to be large-scale. Most of the site's 1,600 staff are expected to become involved in the decommissioning, which will last well into the next century.

The decision to close the plant has been met with disappointment — though not surprise — by Britain's nuclear industry. Dounreay has rarely been out of the headlines in recent years because of its controversial safety record.

Environmental campaigners such as the Friends of the Earth group, on the other hand, are delighted with the decision, but want an immediate end to reprocessing at Dounreay, which is the smaller of Britain's reprocessing facilities, and deals only with fuel from experimental reactors.

The campaigners are also calling on the government to shut down the Thermal Oxide Reprocessing Plant (THORP), at Sellafield in the northwest of England, which handles the bulk of waste from nuclear



Safely grazing? Dounreay's controversial safety record has kept it in the headlines in recent years.

power plants in the United Kingdom and overseas.

The Sellafield plant, however, will probably remain unaffected by the Dounreay decision. THORP remains at the centre of plans for Britain to attract the world's reprocessing business, particularly in the event of a revival in nuclear power as a source of energy that does not contribute to global warming.

Dounreay is being closed because there is no "economic case" for supporting commercial reprocessing, according to the United Kingdom Atomic Energy Authority, which advised the government to close the site.

However, sources in the UK Department of the Environment say that the site's controversial management and safety record, and

the ensuing negative publicity, were also factors in the decision.

The government had been considering Dounreay's future for a number of months. The site is badly contaminated. And it is currently closed while officials from the Environment Agency and the Nuclear Installations Inspectorate carry out an audit of safety procedures following concerns over the security of plutonium at the plant, and the accidental cutting of an electric power cable.

Dounreay is the most complex of Britain's civilian nuclear sites. Commercial reprocessing of waste was just one of its various activities. The site was initially built in 1956 as a pilot plant for Britain's now-defunct fast-breeder reactor programme, which was cancelled in 1988. **Ehsan Masood**

Germany owns up to weapons-grade uranium deal with Russians

[MUNICH] The German federal government, after years of evasion about the likely source of weapons-grade highly enriched uranium (HEU) fuel required by a controversial new research reactor, has admitted negotiating a bilateral agreement with the Russian government for the delivery of such fuel.

According to the deal, which has yet to be approved by the Russian parliament, the Russian company Techsnabexport will provide the German company Nukern with up to 1,200 kg HEU specifically for use in the FRM-II reactor being built by the Technical University of Munich.

Together with a supply of 400 kg HEU already guaranteed from European sources, this represents sufficient fuel for the FRM-II's expected 40-year lifespan.

The United States has protested to the Russian government that the deal could encourage a further increase in commerce in HEU, something it has been attempting to limit through amendments to the 1969 Non-

Proliferation Treaty, for example the Reduced Enrichment for Research and Test Reactors (RERTR) agreement of 1978.

According to the RERTR agreement, the United States — the West's only supplier of HEU until the collapse of the Soviet Union — delivers HEU only to those research reactors prepared to convert eventually to low-enriched uranium (LEU), or those technically unable to do so.

The FRM-II is the only one of 12 large research reactors built since 1978 that has been designed to burn HEU. The United States has already bought up large quantities of Russian HEU stockpiles to prevent it from coming on to the general market.

The US Department of Energy claims that the FRM-II could easily convert to LEU and still deliver the same neutron flux. But the Technical University has aggressively resisted local and international pressure to redesign the reactor, arguing that this would require the university to submit to new time-

consuming licensing procedures.

Gert von Hassel, a spokesman for the FRM-II programme, maintains that the research reactor could still receive HEU fuel from the United States through a US-Euratom agreement signed two years ago which allows delivery of HEU fuel to European reactors that meet RERTR criteria.

But the US energy department has made clear that the FRM-II will not receive US HEU fuel even if its operators were to decide in future to meet criteria by agreeing to convert, because it has refused to convert while in the design stage.

Irene Sturm, a member of the Bavarian parliament for the Green party, claims that the Technical University has "deliberately tried to obscure the facts about negotiations with Russia over delivery of HEU".

But von Hassel rejects these charges. "We have tried to explain technical matters to the general public in a simple way," he says. **Alison Abbott**

Senators seek secure funds for research

[WASHINGTON] Supporters of science in the US Senate are set to introduce a new, long-term proposal for civilian research funding. They hope the measure will have brighter prospects than an existing plan — the National Research Investment Act, or S1305 — which would double such spending in ten years but has attracted the support of only 17 out of 100 senators.

Senators Bill Frist (Republican, Tennessee) and Jay Rockefeller (Democrat, West Virginia) may introduce the proposal as early as this week. It will propose that total funding for non-defence science and technology should grow each year by an amount that exceeds the rate of inflation — possibly by two per cent — and will also attempt to set a 'floor' below which the level of such spending will not fall.

Speaking at a meeting on Capitol Hill last week, Rockefeller repeated his warning that S1305 could not pass (see *Nature* 393, 4; 1998). He said that he and Frist would introduce an alternative that "wouldn't double [spending] but would look at a large amount of money".

The bill would include a basis for spending priorities, and also propose ways of mea-



Rockefeller: 'viable' plan for more spending.

suring the efficiency of research programmes.

Science lobbyists say that the proposal could double the total annual investment of \$34 billion within 12 years, although the figures in it were still being revised late last week. They expect that Senate supporters of the earlier proposal — including Phil Gramm (Republican, Texas), Joseph Lieberman (Democrat, Connecticut), Jeff Bingaman (Democrat, New Mexico) and Pete Domenici (Republican, New Mexico) — will co-sponsor the new proposal.

The new bill will differ from S1305 by giving only one figure for all science agencies,

allowing some agencies to fare better than others. It will also include principles outlining how the money should be allocated, based on a document published last year by the Senate Science and Technology Caucus.

Backers of the proposal hope that the inclusion of these principles, together with the pledge to measure efficiency, will help to attract support for a companion measure in the House of Representatives. Frist and Rockefeller hope to get their measure passed quickly by the Senate Commerce Committee, although the prospects for it passing the full Senate this year do not look bright.

Like S1305, the new proposal is an authorization bill which would not assure annual funding for science, even if it is passed. But scientific societies see such a measure as an important means of expressing bipartisan support for science in the Congress, and influencing future years' budgets.

Bingaman, Lieberman and Rick Santorum (Republican, Pennsylvania) have introduced a similar measure that would authorize basic and applied research spending at the Department of Defense at two per cent above the rate of inflation for the next ten years.

Colin Macilwain

Espionage verdict prompts call for retraction of polymerase paper

[SAN DIEGO] A Californian biotechnology company has been found guilty of stealing research from a former junior laboratory researcher at the University of California at San Diego (UCSD), in a case of scientific espionage.

The researcher, Huguette Pelletier, is now at the Baylor College of Medicine in Houston, Texas. She is trying to achieve the retraction of an article on rat DNA polymerase β — published in the journal *Cell* by researchers at the company in 1994 — because she believes it was based on the apparently pilfered research.

The dispute sets the stage for an intriguing, if rare, dilemma: how to balance the decision of the court with the scientific publication record.

A state court jury ruled last month that La Jolla-based Agouron Pharmaceuticals and a wife-and-husband team of its researchers — Michele A. McTigue and Jay F. Davies II — misappropriated the methodology to grow the crystal for a potentially valuable protein, rat DNA polymerase β . The jury said that Agouron should pay \$200,000 for the theft of trade secrets, and all the legal costs — estimated at \$500,000.

The Agouron article describing the crystallization of the enzyme that is believed to repair DNA was published in *Cell* on

25 March 1994. "I still want the paper retracted — that always has been my main goal," says Pelletier, a crystallographer who published her own more comprehensive article on the DNA polymerase β in June 1994 in the journal *Science*.

Pelletier also continues to maintain that the Agouron scientists cannot scientifically support the research they published in their paper — an allegation that Agouron denies.

Benjamin Lewin, the editor of *Cell*, did not respond to either verbal or written questions submitted last week to the journal's office in Boston.

Repeated requests, beginning in 1994, that Pelletier made to *Cell* to address "scientific discrepancies" in the Agouron article have been rejected.

Despite the jury's decision, Agouron's spokeswoman Donna Nichols insisted that the *Cell* publication was valid research, performed independently from Pelletier's work. Neither Agouron nor its scientists had any plans to retract the article, she said.

Agouron and its scientists were stunned by the verdict, which Nichols rejected as a "ridiculous" decision reached by scientifically unknowledgeable people who "made the wrong decision". Neither of the two Agouron scientists would comment.

The company — best known for developing one of the protease-inhibitor

drugs used against AIDS — has asked the judge to reject the verdict, and is seeking to appeal against the jury's ruling if the judge lets the decision stand.

Although the jury has found that the research was taken from a federally funded laboratory at UCSD, the university played no role in the case, having declined Pelletier's request to join with her in the legal case when it began four years ago. A UCSD spokeswoman declined to comment on the verdict, announced on 27 May.

The scientific theft occurred in the laboratory of Joseph Kraut, now a UCSD professor emeritus of biochemistry. Pelletier, McTigue and Davies all received their doctorates in Kraut's laboratory.

In testimony and court records, Pelletier's San Diego attorney, Daniel B. MacLeod, described McTigue as a scientific "mole who acted as a ferret", secretly seeking the latest techniques discovered by Pelletier in the Kraut laboratory to be transferred to Davies at Agouron.

Agouron had a team of nearly a dozen scientists working on the polymerase β project in 1993–94, when the publicly traded firm injected more than \$1.7 million into an effort to create a 'blockbuster' drug to reduce the side-effects of anticancer medications. But the project proved unsuccessful and was abandoned. **Rex Dalton**

NIH 'should help sharing of research tools'

[WASHINGTON] The US National Institutes of Health (NIH) has been urged to use its full moral and legal authority to promote the freer sharing of research tools between scientists.

The call has come from an expert panel set up last year by Harold Varmus, the director of NIH, to study the problems faced by NIH-funded scientists in obtaining access to research tools — from monoclonal antibodies to cloning tools — from other researchers in universities and industry.

Among the panel's recommendations are that NIH should aggressively assert its right, at the time of making a grant, to take title to a grant recipient's inventions in 'exceptional circumstances'. The panel says that NIH could use this provision of a 1980 technology transfer law, the Bayh-Dole Act, to lay claim to research tools produced for broad use.

The panel also says that NIH could assert its legal 'march-in' rights to allow the agency to force grant recipients who are sitting on patentable inventions to license them on reasonable terms.

In its report, issued last week at a meeting of Varmus's advisory committee, the seven-member Working Group on Research Tools warns that current trends "pose a serious threat to the best interests of biomedical research".

Such problems are familiar to scientists who have tried to obtain research tools from colleagues at other institutions or in industry, and found that their requests frequently elicit long and unfruitful negotiations over terms.

These terms can include a right of approval by the donating institution or company before the recipient submits a manuscript for publication, and delays in publication. Particularly contentious are attempts by the provider to assert rights to ownership or licensing of future discoveries made with the tool.

There are also ubiquitous prohibitions on the requesting scientist's sharing of materials with other researchers or institutions, and restrictions on the tools' use to a particular experiment, among others.

The delays incurred while university technology-transfer offices negotiate terms for their scientists are likely to grow as such agreements proliferate. The University of Pennsylvania, for example, reviewed 425 materials transfer agreements in the last year — an increase of 115 per cent over the preceding year.

The working group is chaired by Rebecca Eisenberg, a professor of law at the University of Michigan in Ann Arbor. She admitted that "there are limits to NIH's authority" to influence current practices, particularly in industry. "But so far NIH hasn't come close to exercising the full extent of its power under current law," she says.



For example, NIH has a legal right to use without charge or permission any patented research tool arising from research that it funds. But in practice some grant recipients have refused to provide the tools. The working group suggests that NIH rewrite its policies to require grant recipients to provide the agency with unique research tools that they develop.

The working group interviewed representatives of biotechnology and pharmaceutical firms, university technology-transfer professionals and some NIH-funded scientists. It found that all three groups agree that there is a problem, but disagree strongly on its nature and on solutions.

The report recommends that NIH set up a Research Tools Forum to bring together representatives from each group to develop a common set of guiding principles for the sharing of tools.

Varmus said that, in line with another recommendation from the working group, he has asked the NIH's Office of Technology Transfer to develop guidelines for institutions that receive NIH funds, explaining what are considered 'reasonable' terms in

the licences and materials-transfer agreements that they negotiate.

These guidelines are to be submitted for discussion to a forum of stakeholders from university technology-transfer offices, industry and laboratories within six months.

Eisenberg said that non-binding guidelines are needed in the light of the "considerable confusion" her group found among university officials about the kinds of terms they should be negotiating.

Universities that have refused to agree to terms they consider unreasonable have felt undermined when competing universities have agreed to the same terms. And many industry representatives said they felt universities were unsure of the nature of their mission with respect to sharing research tools.

The working group also urged NIH to use its moral influence to steer towards practices that are, in legal fact, beyond its control. "NIH can certainly set an example," said Eisenberg.

Eric Lander, director of the Whitehead Institute/Massachusetts Institute of Technology Center for Genome Research, and a member of Varmus's advisory committee, suggested that a university 'cartel' approach to the use of strictly defined research tools might encourage institutions and grant recipients to agree to looser terms for making tools available.

But Eisenberg responded that a cartel would meet practical obstacles, although it was "a great idea in theory". She said that cartels work best with homogeneous products and owners, whereas the tools and owners under discussion vary widely. And universities and industry alike are averse to non-exclusive licensing.

"Nobody is going to be willing to enter into some sort of cross-licensing arrangement that undermines the value of exclusivity for... whatever it is they are selling," she said.

Meredith Wadman

France smooths the way for foreigners

[PARIS] The notorious bureaucratic obstacles facing foreign scientists wishing to work in France may soon be a thing of the past.

Under new measures announced by the government, foreign scientists will in future be considered as a special category, and be able to obtain visas without having first to obtain a work permit — a formal invitation from a recognized research body will suffice.

The measures are intended to encourage greater scientific immigration, which has suffered under France's increasingly strict immigration laws, deterring many prospective visitors (see *Nature* 368, 6; 1994). Under the new rules, employers wishing to invite scientists will also benefit from

simplified procedures, freeing them, for example, from laws requiring that all jobs be advertised nationally to establish that suitable French candidates are unavailable.

Similarly, researchers' families will be awarded visas — and work permits — simultaneously with the approval of the invited scientist's dossier. At present, strict administrative rules often prevent researchers from being joined by their families for up to two years.

The circular setting out the new rules says that an influx of overseas scientists is in the "higher interests of our country", and that applications can only be blocked for national security reasons.

Declan Butler

French researchers reject reform plans...

[PARIS] Researchers at the France's Centre National de la Recherche Scientifique (CNRS), Europe's largest fundamental research agency, have rejected proposals to slash the size of the body that evaluates its laboratories and administers recruitment.

The recommendations that the national committee of the CNRS should be drastically reduced in size were made by Claude Allègre, French minister of national education, research and technology. The committee, which also plays a major role in shaping CNRS's research priorities, comprises about 40 sections covering various disciplines, each made up of scientists from the research agencies and universities.

Allègre wants to halve the number of sections as part of his plans to reduce bureaucracy in French science (see *Nature* 388, 7; 1997).

But many scientists have argued that a much reduced national committee would lack the manpower and expertise to carry out evaluation properly.

A recent report commissioned by CNRS from Jean Pailhous, a CNRS researcher at the Université Aix-Marseille 2, also recommended that the sections only be reduced to 36 (see *Nature* 392, 8; 1998). This view gains strong support from the results of a three-month consultation at CNRS, which is due to be released shortly.

It shows that three-quarters of all CNRS laboratory directors oppose any large reduction in the number of sections in the committee, with only 10 per cent supporting such a move. Similarly, 81 per cent oppose a reduction in the number of members in each section.

A narrow majority — 54 per cent — agreed that CNRS's seven departments should play a larger role in running the agency. A further 30 per cent who opposed this said they would accept it if assurances were given that the departments would base their strategic thinking on that of the national committee.

The result confirms the profound attach-

ment of CNRS researchers to the national committee, which is unique in that two-thirds of its members are elected directly by the scientific community. Indeed, most researchers want the current balance of elected scientists to be maintained, with only 2 per cent of those questioned supporting a reduction of this proportion to less than 50 per cent.

One member of the CNRS administrative council predicts that, given this broad resistance, Allègre is unlikely to be able to proceed with his proposed reforms. A clearer idea of what, if any, reforms of the committee are envisaged should emerge in two weeks' time when the CNRS administrative council itself is due to propose a series of reforms for the agency.

The main change is likely to be a reinforcement of the role of the administrative council in strategic planning, in particular by

being involved from the outset in preparing reforms and decisions.

Although the board votes on the budget and other strategic issues, at present its role is largely restricted to rubber-stamping proposals from the director-general. But Allègre is keen that administrative councils should play a real role in setting strategy at all the research agencies, including the medical research agency INSERM (see below), and that the director-generals should be more concerned with day-to-day management.

Meanwhile, the scientific trade unions, having gained the scientific community's backing in their opposition to Allègre's proposals to dismantle the national committee, now seem to be supporting other changes to reduce bureaucracy, increase international input in evaluation and encourage multidisciplinary.

Declan Butler

...as medical agency agrees to a compromise

[PARIS] Extensive negotiations over reform at the French national biomedical research agency have ended with the adoption of a text reflecting last-minute concessions to trade unions representing the agency's researchers. The planned reform to the statutes of the agency, INSERM, is intended to streamline its internal decision-making.

The trade unions say they remain unhappy about the final text adopted by the agency's administrative council, which they claim will increase political control over INSERM (see *Nature* 391, 110; 1998) by increasing the power of this council, but that the last-minute

concessions won have eased their major concerns.

The government had already abandoned the most controversial elements of the decree — such as splitting INSERM into five departments — following opposition from researchers.

But the modified reform text was nonetheless rejected by staff union representatives last month. They claimed management had given insufficient assurances on the scope of new powers given to the administrative council to make strategic decisions on research directions and funding.

Such decisions were previously reached through

agreement by the scientific community and INSERM management, and the unions argued that assurances were needed over the continued role of INSERM's scientific commissions in decision-making.

The administrative council has bowed to these demands, and the adopted text restricts the council's role to deciding on the 'broad lines' of policy, staff and budgetary decisions. Also, the scientific committees must be consulted not only on the creation, modification or abolition of programmes, but also on the financing of individual INSERM laboratories.

D.B.

Harvard's 'oncomouse' fails to win Canadian patent

[MONTREAL] A Canadian federal court judge has ruled that Harvard University's 'oncomouse', which was genetically engineered to be susceptible to cancers, and the rights to which are now owned by DuPont, is not patentable in Canada.

After receiving a US patent on both the biotechnology process used to 'create' the oncomouse and on the mouse itself, Harvard's president and university fellows had applied for a Canadian patent — the

first such patent sought in this country for a mammal.

The patent office had accepted the claim for the genetic engineering process, but it rejected the claim on the mammal itself. Harvard appealed to Canada's federal court. But Judge Marc Nadon has dismissed the appeal, saying: "They have created a method to inject eggs with a *myc* gene but they have not invented the mouse."

Nadon said that parliament could alter legislation so that mammals can be patented

if it wishes. But he said that he was not prepared to manipulate the meaning of the words of the Patent Act to the degree currently necessary to do so.

Michelle Swenarchuck, counsel and director of international programmes at the Canadian Environmental Law Association, says: "Justice Nadon has correctly placed this matter at the door of elected officials." Harvard is expected to appeal against the judge's ruling, taking the issue to Canada's supreme court if necessary. David Spurgeon

Swiss reject curbs on genetic engineering

[MUNICH] Proposals to severely restrict genetic engineering in Switzerland — for example by banning the use of transgenic animals in research and halting field trials with transgenic crops — were rejected in a referendum last Sunday (7 June) by a surprisingly strong majority.

Two-thirds of the voters rejected restrictions proposed in 1992 by 70 pressure groups opposed to biotechnology, bringing to an end months of bitter — and expensive — campaigning by both sides. The turnout of just over 40 per cent was particularly high for Switzerland, under whose system of government the electorate is frequently asked to vote on relatively minor legislative matters.

More than 80 per cent of the French- and Italian-speaking parts of the country voted 'no' in the referendum. But the 'no' vote was also higher than expected in German-speaking areas, where the population is more mistrustful of genetic engineering. The referendum's proposals were rejected in all of Switzerland's 26 cantons.

Most recent surveys had indicated that the vote would be very close and could swing either way (see *Nature* 393, 405; 1998). "We were sitting in our department all Sunday afternoon, listening to the radio and biting

our nails," says Sebastian Brandner, an assistant professor at the Institute for Neuropathology at the University of Zürich. "When the result became clear, the relief among all colleagues was palpable."

Many believe that the willingness of prominent scientists to write popular newspaper columns and appear on television contributed to the clear 'no' vote. Those who had put themselves in the media spotlight included the Nobel laureate Rolf Zinkernagel, head of immunology at the University of Zürich, and Adriano Aguzzi, head of neuropathology at the university.

Politicians and researchers had warned that restrictions on genetic engineering would seriously harm the Swiss economy, and could cause the loss of tens of thousands of jobs in academic and industrial research departments.

The Swiss government will now continue with an initiative of its own called 'Gen-Lex', launched last year, which was intended in part to encourage waivers to vote 'no' in the referendum by tightening and coordinating existing regulations on genetic engineering (see *Nature* 391, 312; 1998).

The Gen-Lex initiative has broad support, including backing from industry. It

includes, for example, a proposed adaptation of the animal protection law to ban the breeding of transgenic farm animals that might suffer pain.

It also proposes to extend the liability period for all genetic engineering activities which is 20 years in the European Union, to 30 years in Switzerland. This means that a company or a research institute will be liable to pay for damage caused by a genetically engineered organism for up to 30 years after its first approved release into the environment.

These rules will come into effect in 1999 at the earliest. But a national ethics committee — another Gen-Lex initiative — has already been set up as a permanent advisory body for the government. This committee of 12 experts in science, ethics and law, advises on all ethical issues pertinent to the use of animals in genetics research.

The committee will also serve as an "institutionalized dialogue platform between experts, the government and the public," according to Conrad Engler, head of gene technology at Interpharma, the association of the Swiss pharmaceutical industry.

Researchers recognize lack of dialogue as one reason for public scepticism about biotechnology.

Quirin Schiermeier

India boosts budget for atomic research in wake of bomb tests

[NEW DELHI] India, which carried out nuclear tests and subsequently declared itself a nuclear weapons state last month, has increased the budget of its Department of Atomic Energy (DAE) by 30 per cent. The rise is the first major hike in the department's 50-year history.

But officials are still insisting that the enhanced funding, which takes the department's budget to US\$652 million, is intended to lead to the generation of more electricity by nuclear power, rather than to building improved nuclear bombs.

Overall, the 1998–99 budget presented to the country's parliament last week allocates \$2,443 million for non-capital spending on science and technology, 22 per cent more than last year. Expenditure on space will rise by 52 per cent, to \$401 million.

The allocation for defence research has gone up by 24 per cent, to \$619 million. The move has inevitably raised speculation that the Bharatiya Janata Party (BJP) has decided to maintain the momentum of the nuclear blast by strengthening all the three strategic departments associated with nuclear weapons.

But Rajagopalan Chidambaram, DAE secretary and a key scientist in the nuclear testing, denies this. "The decision on budget allocation was taken well before the Pokhran

Area	Amount	Percentage increase 98/99
Atomic energy	652	30.4%
Space	401	51.9%
Defence research	619	24%
Agricultural research	252	58%
Environmental science	205	51.9%
Industrial research	160	6%
Science and technology	156	2%
Medical research	129	19.4%
Non-conventional energy	102	108.2%
Electronics	53	29.2%
Biotechnology	29	20.8%
Oceanography	27	3.8%
Total	2,785	

tests," he says. "The main increase is in the plan allocation for nuclear power."

Chidambaram points out that, while DAE's research facilities receive only a 40 per cent boost in funding, the nuclear power schemes get an 86 per cent rise. "The government is very keen to encourage the power sector," he says.

Krishnagopala Iyengar, the former head of DAE, describes the Indian weapons programme as a low-cost affair. He says work related to the nuclear bomb over many years did not even have a regular budget, and was a "part-time" job for DAE

scientists. Plutonium was a by-product of the nuclear energy programme and, according to Iyengar, the five bombs tested by India cost just \$1.25 million to fabricate.

DAE plans to use the enhanced funding for the development of intense particle beams for research, and for an advanced heavy-water reactor to burn thorium — abundant in India.

A spokesman for the BJP government said that much of the new space budget is intended to build the next-generation INSAT-3 telecommunication satellites, which will replace the INSAT-2E series.

Rocket development also gets a significant share, and officials say the objective is to develop the Geostationary Launch Vehicle for the INSAT-3 satellites. The budget provides \$12.5 million for a new launch pad for the vehicle.

The research community in general has welcomed the extra funding for science. Spending on education has increased by 50 per cent to \$1,762 million, and \$6.25 million has been set aside for improving research infrastructure in universities. "This is the first time a new initiative has been taken to rebuild science and engineering departments in universities," says Valangiman Ramamurthy, secretary in the science ministry.

K. S. Jayaraman

Senate rejects plea for greenhouse-gas research funds

[WASHINGTON] Senate appropriators have ignored a Department of Energy request for additional research funding to help the United States reduce its greenhouse gas emissions. But they have agreed to fully fund the department's scientific programmes.

The Senate's energy and water appropriations subcommittee, chaired by Pete Domenici (Republican, New Mexico), rejected increases in solar and geothermal energy research that had been requested under the administration's Climate Change Technology Initiative (see *Nature* 391, 619; 1998).

But the panel supported the modest increases requested for the science programmes, which support most physics research in the United States, and for the construction of the Advanced Neutron Spallation Source at Oak Ridge National Laboratory in Tennessee.

Modest additional funding for renewable energy research is the only significant action that the Clinton administration has proposed to cut greenhouse gas emissions since it signed the Kyoto protocol last November, and its rejection by the Senate is a serious blow to any compromise between the administration and opponents of the accord.

Vice-President Al Gore said on Monday (8 June) that he would keep fighting for the initiative, citing new data that global temperatures reached record average highs in the first five months of the year.

Prize for creator of the World Wide Web

[MUNICH] Tim Berners-Lee, creator of the World Wide Web, has received this year's DM200,000 (US\$112,700) Eduard Rhein Foundation prize for technology, one of the most valuable such prizes in Europe. Berners-Lee created the World Wide Web while working at the European Particle Physics Laboratory (CERN) in Geneva.

The Eduard Rhein Foundation prize for basic research was awarded to Jacob Ziv, president of the Israel Academy of Sciences, for his pioneering work on information and coding theories.

Rhein was a German writer and inventor who died in 1993.

Prusiner was turned down for UK BSE grant

[LONDON] Two of Britain's main food research agencies rejected in 1991 an application for a grant, to assess the risks to humans of eating beef infected with bovine spongiform encephalopathy (BSE), from

Stanley Prusiner, the winner of the 1997 Nobel prize for his work on prions. He is professor of neurology and biochemistry at the University of California, San Francisco.

Prusiner wanted to use transgenic mouse models to assess the risk of humans contracting a fatal prion disease from eating BSE-infected beef. He applied with a British collaborator — Gareth Roberts — to the agriculture ministry and to the Agriculture and Food Research Council in 1991, when the UK government believed that beef was safe to eat.

A second application, in 1996, was also rejected. "I would argue that this was a major mistake," Prusiner told the public inquiry into BSE last week. "It is not as though we gave up."

NASA retains manager for Hubble's successor

[WASHINGTON] The US Space Telescope Science Institute last week got a new lease on the future. The US space agency NASA has selected the Baltimore, Maryland-based institute to manage science operations for the Next Generation Space Telescope (NGST) proposed for launch in 2007. The institute currently manages the Hubble Space Telescope from the campus of Johns Hopkins University under contract to the Association of Universities for Research in Astronomy.

The European Space Agency recently began its own study of the NGST project, which is expected to involve substantial European participation. An international workshop on scientific priorities for the NGST will be held next week in Liège, Belgium.

Euro parliament stands firm on Framework cash

[MUNICH] The European Parliament's research committee has rejected the European Council's proposal for financing the European Union's fifth five-year Framework programme of research (FP5) to the tune of ECU14 billion (US\$15.5 billion). As expected, the committee is holding out for ECU16.3 billion, the figure originally proposed by the commission, and a small reduction to the sum parliament had proposed in its first reading.

The full parliamentary plenary session is almost certain to approve this sum in its second reading of the Framework programme proposal next week, and the issue will then go to arbitration.

The conciliation procedure will also have to resolve a second disagreement over the financing of FP5 — Spain's refusal to vote on issues involving financing after 1999 as part of its battle to persuade the European Union to maintain its subsidies to Spain after this date.

Neutrinos are not as light as light

[LONDON] The Super-Kamiokande experiment in Japan has measured a non-zero mass for the neutrino. The announcement, made on 4 June, was based on two years of data that reveal an 'oscillation' of one type of neutrino into another. This is only possible if neutrinos — unlike photons, the particles of light — have some mass. A preliminary detection was made months ago (see *Nature* 391, 123–124; 1998), but the team is now more sure of its results.

The discovery of oscillations could solve the 'missing solar neutrino' problem, and means that neutrinos form some part of the 'dark matter' that must be present in the Universe. But, as the standard model of particle physics predicts that the neutrino has zero mass, it appears that a new fundamental theory of physics is needed.

Judge fails to see the joke in improbable feud

[BOSTON] A judge in the Illinois federal district court has dismissed the complaint by the publisher of one science humour magazine against the editor of another. George Scherr, publisher of the *Journal of Irreproducible Results*, had accused Marc Abrahams, editor of rival magazine the *Annals of Improbable Research* (see *Nature* 389, 431; 1997), of libel, racketeering, "massive perversion and fraud" and other improprieties. But the court ruled that these alleged offences were outside the jurisdiction of Illinois.

Scherr can appeal against the decision or file lawsuits in other states. But lawyers familiar with the case say the judge's ruling will make it harder, and more expensive, for him to pursue his grievance. Meanwhile, Scherr is pressing forward with his campaign for ownership of the term "Ig Nobel Prize" — a claim that has been rejected by the US Patent and Trademark Office, and subsequently appealed against.

Time runs out for historic astronomical instrument

[LONDON] An unknown seventeenth-century designer of astronomical instruments is believed to be among the first to put the 'millennium bug' into a piece of equipment. His — or her — identity is not known. But their brass Equatorium instrument, housed in the Liverpool Museum, in northwest England, will stop working at midnight on 31 December 1999. "It is a little sad that the working life of this 400-year-old, unique instrument comes to a close in 18 months," says Martin Suggett, curator of Earth and physical sciences at the museum.

British Biotech responds to allegations

Sir — Your News article on British Biotech continued to give credence to allegations which the company's board has repeatedly stated are "without substance or represent purely personal opinions of Dr [Andrew] Millar" (*Nature* 393, 299; 1998).

To provide a detailed and factual basis for this statement, the company published a 32-page circular to shareholders on 19 May. The circular was compiled with substantial due diligence and full legal verification of the information it contained. It was issued with the authority of the board of directors. In the circular, John Raisman, British Biotech's chairman, stated that, following extensive reviews, the board is satisfied "that the company has throughout acted in good faith in the interests of its shareholders and without any impropriety".

Against this background the company remains concerned that *Nature* may have given undue prominence to views that British Biotech directors traded in the company's shares in 1995 when they should not have done, given the status of the clinical trials of batimastat, a potential anti-cancer drug; or that they did not inform shareholders soon enough of the negative outcome of the regulatory review by the European Medicines Evaluation Agency (EMA) of Zacutex, a potential new treatment for acute pancreatitis. These are serious allegations and they were given detailed analysis in the circular.

The facts with batimastat are clear. This injectable matrix metalloproteinase inhibitor commenced a phase III randomized controlled clinical trial in late 1994 in patients with advanced inoperable abdominal cancer. All the patients had malignant ascites, an accumulation of cancer fluid in the abdomen, and all were within months of expected death. 'Serious adverse events', the name given to any severe medical problem, were reported in these patients commonly but, importantly, at the time of the share transactions in

early 1995, there was no evidence that serious adverse events were occurring more frequently in batimastat-treated patients than in the control patients not receiving the drug. Furthermore, in a previous phase II study in a similar group of patients with advanced ascites derived from ovarian cancer, batimastat had shown clinical benefits and been reasonably well tolerated.

There was therefore no reason to expect that the phase III trial would not continue as planned. In fact, as late as one week before the share transactions cited by Millar, a meeting of the London Gynaecology Oncology Group, comprised of cancer specialists undertaking the trial, concluded simply that changes to the dosing of spironolactone, a diuretic, may be necessary and that the study would be discussed again in three months' time.

As a result of reviewing these events in detail the board stated clearly in the circular that it "is satisfied there is no substance to the allegations that certain of the then directors of the company dealt in shares when they should not have done".

In relation to whether British Biotech made adequate disclosure of the progress of its marketing authorization application through the EMA review process, the circular gives a detailed chronology of events and shows that "all public statements in relation to Zacutex were factually accurate and reflected the company's reasonable expectations of the prospects for Zacutex at the time they were made". No company can aspire to higher standards of disclosure than this.

In your editorial of 28 May (*Nature* 393, 291; 1998) you thank Millar for highlighting the ease with which the boundary between hope and hype can break down, with the implication that this situation applies to Zacutex. However, a large-scale (1,500-patient) international trial of Zacutex in acute pancreatitis is still under way. The outcome of this double-

blind pivotal study, expected at the end of 1998, cannot be predicted at this stage. It is designed to confirm whether Zacutex, a platelet activating factor antagonist, can reduce mortality in the treatment of acute pancreatitis as suggested by a previous UK phase III study. If successful, Zacutex may represent a significant breakthrough in the treatment of this disease.

However, as a result of Millar's repeated unauthorized unblinding of the data being collected in this study, in contravention of good clinical practice guidelines and all normal company and regulatory agency procedures, the integrity of this study may have been jeopardized such that it cannot be used for regulatory approval of the drug, even if the results of the trial are positive. British Biotech recently commissioned an external regulatory audit of all double-blind trials conducted by the company under Millar's supervision during his period at the company. The audit findings have shown that not only the Zacutex study but also one of the company's pivotal trials with marimastat, a potentially important new anti-cancer agent, may have been damaged in this way. Such unprofessional conduct by a senior medical executive would more normally attract condemnation rather than thanks.

It is regrettable that the events of the last few months at British Biotech have damaged the reputation and value of one of Europe's leading biotech businesses. Biotech companies are uniquely vulnerable to investor sentiment irrespective of whether this is based on scientific data or unsubstantiated rumour.

Keith McCullagh (Chief Executive)
British Biotech, Watlington Road,
Oxford OX4 5LY, UK.

When writing the story, many attempts were made to contact McCullagh, via the company switchboard and through British Biotech's public relations officer. The calls were not returned. — Editor, *Nature*.

Way forward at NSF

Sir — A recent News story by Colin Macilwain implies that the US Government Performance and Results Act (GPRA) will prompt National Science Foundation (NSF) staff to look over the shoulders of grant recipients in an unprecedented way (*Nature* 393, 100; 1998). The foundation has always required grant recipients to report on annual progress and to provide final project reports. Rather than monitoring grants more intensely, we will simply be capturing

this information in a way more useful for reporting on performance of our programmes.

We are also asking for the help of grant recipients in identifying less immediately obvious effects of their activities, and to communicate those in terms that lay readers can understand.

For GPRA, NSF will examine and report on the performance of programmes as a whole, rather than the performance of individual grants. Processes are already in place in some directorates to obtain

summary information on programme activities. GPRA requires that we obtain that information in a way that can be aggregated to the level of NSF as a whole and prepared for assessment by panels of external experts. We are attempting to structure this aspect of GPRA so that it has minimal impact on NSF staff and the community.

Judy Sunley (Assistant to the Director for Science Policy and Planning)
National Science Foundation, 4201 Wilson
Boulevard, Arlington, Virginia 22230, USA
e-mail: jsunley@nsf.gov

Ethical discourse by science-in-fiction

It can be difficult to discuss ethical dilemmas in the academic environment. One way of doing it is through 'science-in-fiction'. The following 'science renga' shows how this is done with virtually total anonymity.

Carl Djerassi

"Ethical discourse through science-in-fiction" is the formal title of Medicine 256, a graduate course I have recently offered under the auspices of the Center for Biomedical Ethics of Stanford University's School of Medicine in California. Given the sudden proliferation in the United States of new courses on biomedical ethics, what is different about Medicine 256?

Scientists operate within a tribal culture, the rules, mores and idiosyncrasies of which are generally acquired through intellectual osmosis in a mentor-disciple relationship. Scientific 'street smarts' are absorbed by observing the mentor's self-interested concerns with publication practices and priorities, the order of the authors, the choice of journal, the striving for academic tenure, grantsmanship and even the Nobel prize. On their own, disciples discover the 'glass ceiling' for women in a male-dominated enterprise, the inherent collegiality of scientific research, and also the *Schadenfreude* generated by brutal competition. Most of these issues are related to the desire for personal recognition and even financial rewards, and each is coloured by ethical nuances.

An effective medium for illuminating such topics is the rarely used literary genre of 'science-in-fiction' (not to be confused with science fiction), in which all aspects of scientific behaviour and scientific facts are described accurately and plausibly. By disguising them in the cloak of fiction, science-in-fiction allows the illustration and discussion of ethical dilemmas that are frequently not raised for reasons of discretion, embarrassment, or fear of retribution.

As an experiment in using science-in-fiction as a didactic tool, I asked 14 graduate students and postdoctoral professionals from 12 different departments at Stanford to compose a short story of up to 10 pages, dealing with the ethical issues associated with relevant behavioural practices in science or medicine.

After I had discussed each with its author and it had been subsequently revised, the stories were distributed among all participants without the authors being identified, to allow unrestricted discourse. The rest of the course was made up of some in-depth and heated discussions of the ethical or behavioural problems raised by the stories.

As well as creating a forum for open discussion and debate, the course also addressed the question of how scientists might communicate better with their col-

leagues and the general public. This led to an attempt to use the Japanese form of renga (linked verse in which stanzas are composed by two or more poets in alternating sequence, often as a form of competition) to create a short story dealing with a scientific ethical dilemma. Each paragraph was composed by a different student who did not know the identity of any previous author. Each student then added a fifteenth paragraph to the 14-paragraph 'science renga', generating 14 new endings. The 'winner' was selected by closed ballot and lightly edited for length. Although it bears the names of all authors — a feature common among scientific papers but virtually unheard in literary publication — none knows who contributed which segment.

Japanese renga bears a resemblance to the process of scientific co-authorship, as it has collegial and competitive aspects, but our experiment is a 'purer' collaboration because each author is associated with the whole enterprise but no identifiable individual component. Further similarity to a scientific paper was pursued by one of the participants, E. Weber Hoen, who composed an abstract of the original 14-paragraph 'science renga' in the form of a 14-line sonnet called "Old goat".

Carl Djerassi is in the Department of Chemistry, Stanford University, Stanford, California 94305-5080, USA.

e-mail: djerassi@stanford.edu

For further discussion of science-in-fiction see <http://www.djerassi.com>

"Old Goat" by E. Weber Hoen

**Old goat, the kid who challenged you is dead
and the rain has clapped its hands of the affair.
And yet, your horny crown is filled with dread,
and your face exaggerates the weight of hair.
Suspicious of the genius in the stones,
you make a science of your steps, and climb,
negotiating past those ghostly bones,**

**which keep from you their insights into time.
It is height you desire, and with that, truth,
to shake your beard on an eternal view,
as if from there you might behold your youth.
The rain, though, has you blind. Below, like you,
the young conspire in fear against their king.
Goat, you are old. You have not learned a thing.**



MARK DOBSON



MARK DOBSON

A science renga

Alfred N. Aldston Jr*, **Dina L. G. Borzekowski¹**, **Jonathan A. Eisen²**, **Sheri L. Fink³**, **E. Weber Hoen⁴**, **Dean Y. Hung⁵**, **Shirley Lin⁶**, **Cynthia T. M. H. Nguyen⁷**, **Julie E. Phillips⁵**, **Michelle Stohlmeyer⁶**, **Cenk Sumen⁸**, **Craig A. Swanson⁹**, **Noriko Takiguchi¹⁰**, **Yvonne Thorstenson¹¹** & **Harriet A. Washington¹²**

"Pall bearing is a young man's job," thought Mankiewitz as Dr Carlson's corner of the casket slipped lower. Mankiewitz was surprised by his advisor's struggle. The hairy billy goat had never appeared to lack vigour. But today, rain dripping from his clumsily cropped beard and the white tufts of hair on his wrists, he seemed a frailer, more tamed version of himself.

The death of Alan Stilwell had shocked Carlson. Yet he could not have hoped for a tidier solution. Alan had been his best postdoc, and a good friend. But when Alan stole his idea their friendship unravelled into bitter competition. Thoughts of Alan's betrayal and its ironic, unpredictable consequences strengthened Carlson.

"Angry at the fates, Professor Carlson? Or just frustrated with their whimsical nature?" Mankiewitz disrupted the silence as the opposite corner of the casket rose. "I always told him not to drive his Miata so fast." Carlson checked his stride, "I guess he'll live

through his publications." As they returned to their positions for the eulogy, Carlson leaned toward Mankiewitz, "Come to my office tomorrow so we can go over the last few weeks of Alan's notes. You were closest to his work." Mankiewitz steadied himself. "Professor," he whispered back, "screw yourself."

After the service, Carlson, Mankiewitz, and a few others lingered by the grave before trudging to their cars. Carlson noticed Mankiewitz entering a limousine with Jesse Stilwell, Alan Stilwell's father. Jesse was a highly decorated member of the National Academy of Sciences, a superstar. It was Jesse who had persuaded Carlson to take Alan into his lab.

As Carlson drove towards his office, his thoughts were paced by the wipers' frenetic beat: Man-kie-witz, Stil-well, Man-. He wondered why Jesse had offered his limo to Mankiewitz. He chided himself for worry-

The death of Alan Stilwell had shocked Carlson. Alan had been his best postdoc. But when Alan stole his idea their friendship unravelled into bitter competition.

ing. Jesse was a trusted colleague, his recent call had assured him of that. Carlson approached his office ready to use his key when he noticed the door was cracked open. His computer screen glowed, outlining the head and neck of the body sitting before it.

This apparition reminded Carlson of his last image of Alan. He had been working with his mouse, deftly clicking his way through the Cambridge Crystallographic Database. A tri-folded smartly headed letter stuck out of his back pocket. Carlson neither would have noticed the letter, nor known what it contained had his editor friend at *Science* not just called. Indeed, the rumours were true. Eldorex Corporation's scientists had just confirmed that untransformed, normal cells could live indefinitely if they expressed telomerase. As Carlson had predicted, senescence could be surpassed. The call also confirmed another of his suspicions that had grown out of Alan's persistent questions about inventors' rights, the seemingly irrelevant articles about the intercalation of hydrophobic and hydrophilic polymers into cell membranes piled on his desk, and those inaccessible conversations with Mankiewitz. The call was the final piece explaining that Alan was neatly bypassing his advisor and the university by licensing to Eldorex the technology necessary to deliver telomerase into specific cells in the body. Carlson thrust the door open.

The person sitting at the computer jumped up and grabbed a small tote bag. Racing out of the office, he shoved Carlson aside. As he grabbed the doorjamb for support, Carlson registered the man's face, especially his hazel eyes; they would not be forgotten. Carlson watched in disbelief as the interloper scrambled down the hall. He thought first to call the police, but embarrassment and fear dissuaded him. He assessed the office. Disks were strewn about, and on the screen were files containing data from Alan's polymer experiments. He wondered how the intruder knew to look on his computer, not Alan's.

Mankiewitz rode in the limousine uneasily, thoughts of the cold cemetery affixed in his mind. A day like today makes one consider cremation, he thought, and shuddered. Alan's silent father sat on the side-facing seat to his left, staring out the window. The day after the accident, Mankiewitz had received a call from him. His manner had been abrupt but not unlike what Mankiewitz expected from busy scientists. "This is Jesse Stilwell," he had said, "Tell me what you know about Alan and the Eldorex Corporation." Too shocked to deny his knowledge, Mankiewitz had simply hedged, "I would feel more comfortable speaking with you in person."

When he entered grad school a dozen years ago, Mankiewitz had piously denied the motivation of money. But as friends purchased homes, cars, and fancy vacations, his ideals tarnished. Now wiser and less idealistic, he stayed in academic science only to network and discover potentially profitable biotech ideas. The other graduate students and postdocs knew Mankiewitz was older, but never questioned his work. He was good at deflecting their curiosity by focusing on their interests. When the younger Stilwell began speaking to him about telomerase, Mankiewitz knew he had found gold. With effort he had convinced Alan to ignore the university stipulations that all discoveries and a portion of any patent royalties were considered University property. Mankiewitz arranged Eldorex's licensing agreement for Alan, not disclosing his finder's fee. He made sure Carlson and the University remained unaware of the arrangements and maintained the pose of a hard-working postdoc. But as the surreptitious project neared completion, Alan had expressed concern and fear.

When he entered grad school, Mankiewitz had piously denied the motivation of money.... Now wiser, he stayed in academic science only to network and discover potentially profitable biotech ideas.

Now that Alan could never proclaim these fears again, Mankiewitz wondered how much and what to reveal to Alan's father.

Jesse Stilwell began the conversation. "I'm the one who sent him to Carlson's lab to work on telomerase. Yet my own son avoided me after he became involved with Eldorex." Mankiewitz had long ago learned that lies were dangerous; one had to remember exactly what had been said. Partial truths were easier; one only said what one wanted to say. "The original idea was Carlson's, and Alan was excited," Mankiewitz responded slowly. "He felt it could be a blockbuster. But he wanted to get the technology to the public as soon as possible." He gauged Jesse Stilwell's reaction carefully. "Alan felt Professor Carlson was stalling, so he went directly to Eldorex. He realized this was dangerous, but felt he had no other options. And I assisted in the agreement with Eldorex." Stilwell looked out the window. "I knew that someone close to Alan must have lured him away from me. I thought it had been Carlson." He paused. "I prefer to learn that it was you."

Carlson sat at his desk, staring at the computer, reliving the bitter experience of having lost telomerase to outside licensing. He had suspected Alan's industrial connection, but had not been willing to confront him. Instead, he had capitalized on Alan's ignorant use of lab computers to collect Alan's data on his own computer in an attempt to reclaim his original idea. Alan's death had given him the impetus he needed to absorb that material. Before the intrusion he was slow, too distracted to be productive. But as he urgently reviewed the data, he began to grasp the extent of his ignorance of Alan's work.

A few days after the funeral, the phone startled him. A woman's voice said briskly. "I'm May Turner, vice-president of one of Eldorex's competitors. You don't know me or the company. Dr Jesse Stilwell, our principal scientific advisor, thought you might be willing to meet... and soon."

They met that evening. Instead of the expected Stilwell, the burglar who had rifled his files accompanied Ms Turner. They were not introduced. Carlson offered his hand to the tall, lithe woman. "Let me get right to the point." Her eyes widened as she tendered a quick smile. "As a lawyer, I'm an observer of human nature. It's been twenty years since your seminal work on cell death was anticipated by a similar publication in *Nature*. I suspect you're disappointed not having received the rewards you expected from your career. I can't return two decades of frustration but I can offer a degree of financial security that academia cannot match." Turner looked him fully in the face. "Alan Stilwell's unfortunate death had for us the advantageous effect of interrupting his work for another lab, work rightfully covered by the university contract. We want you to help us develop that work. This move breaks the possibly incriminating chain of custody, as you have no prior knowledge". "It was my idea," blurted Carlson. "Of course," she said, "but every lawyer knows that reduction to practice is as important as conception."

Soon, it became clear to Carlson that Turner and the intruder had reached the same conclusion as he: the treasures of Alan's work were not revealed in the computer data, but were hidden in the procedural details — the cornerstone of experimental and financial viability. They agreed to meet again. Carlson returned to the lab, to Alan Stilwell's bay. He began scanning through his student's ordered notebooks, flipping the pages. Pictures, gels, text... one blur after the other. Procedure, Results, Ideas, Brainstorm... the labels spun by. He searched for keywords: intercalating, polymer, hydrophobic, telomerase, senescence, cell cycle, cycle. Nothing. As the details blurred, he recalled the police accident report: "Personal items were recovered... but not identifiable because of fire damage... cause unknown."

"You won't find it there." Startled from his whirlwind of thought, Carlson spun toward the door. Mankiewitz stood there, smiling tensely. "I've already looked," he said. "Even before his death. But I've come with a proposition. May I sit down?" He pointed to the lab stool. "Mankiewitz," growled the professor, "I quote: 'screw yourself'." □

¹Center for Research in Disease Prevention,

²Department of Biological Sciences, ³Program in Neurosciences, ⁴Department of Applied Physics,

⁵Department of Genetics, ⁶Department of

Chemistry, ⁷Department of Psychiatry and

Behavioral Sciences, ⁸Department of Microbiology

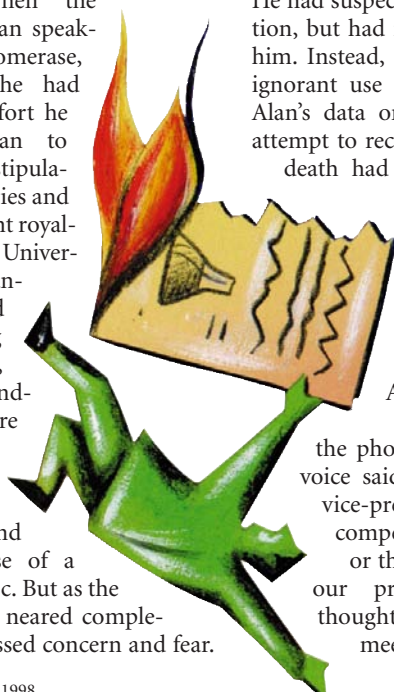
and Immunology, ⁹Department of Pathology,

¹⁰Department of Computer Science, ¹¹Department of

Biochemistry, and ¹²Knight Journalism Fellowship

Program, Stanford University, Stanford, California 94305, USA.

*Deceased



Blueprint for the white plague

Douglas B. Young

With annual deaths from *Mycobacterium tuberculosis* estimated at around three million, this single pathogen claims more human lives than any other. What we now learn from the sequence of its genome should help in devising new strategies to fight it.

“Now — under the microscope the structures of the animal tissues... are brown, while the tubercle bacteria are a beautiful blue.”

More than a century after Robert Koch's historic lecture¹, we are presented with a new perspective on *Mycobacterium tuberculosis* — the complete sequence of its genome, reported by Stewart Cole and his colleagues² on page 537 of this issue. Although no longer blue, it retains a sense of beauty, and the sequence marks a new phase in the battle against one of mankind's most successful predators.

In the final frantic years of the twentieth century, with newly emerging pathogens adapting to a lifestyle dictated by the time it takes to cook a hamburger, it is remarkable that it is the tubercle bacillus — a ponderous microbe more often associated with the nineteenth century — that still claims the title of “captain of all the men of death”. *Mycobacterium tuberculosis* opted for residence in the inner recesses of the human lung probably around the time that cattle were first domesticated, some 10,000 years ago³. It took a long view of life, eschewing the rapid, go-for-growth strategies of more acute pathogens in favour of a leisurely cycle restricted to a single daily round of cell division even in optimal culture conditions. For the most part, it adopted a pattern of peaceful co-existence with its human host in the form of a quiescent or dormant infection, establishing a massive reservoir of infected individuals — possibly including as much as one-third of the world's population. But it can also trigger progressive and deadly destruction of the lungs in the unfortunate minority who suffer from clinical disease (Fig. 1) and act as the source for further dissemination.

Over subsequent millennia *M. tuberculosis* has taken a close professional interest in the evolution of human society. It has exploited opportunities to spread during periods of urbanization and social upheaval, retreating to the margins of society in times of affluence and improved hygiene. The tubercle bacillus has also weathered the antibiotic revolution, in part through mankind's natural reluctance to comply with the requisite six-month intensive

course of treatment, and, more recently, by the development of drug-resistant forms. Moreover, it has forged a deadly partnership with the human immunodeficiency virus (HIV) — increased susceptibility to tuberculosis is associated with the early stages of HIV infection, and tuberculosis in turn accelerates the progression to AIDS.

This history, together with the clues to conquer the tubercle bacillus, is written in its genome. Thanks to Cole *et al.*², we now have the sequence of every potential drug target and of every antigen we may wish to include in a vaccine. But can we convert this mass of information into a useful understanding? It will not be a simple task. It is a good-sized genome — with 4.4 million base pairs, it is surpassed in size only by *Escherichia coli* in the list of bacterial genomes completed to date. It has the metabolic potential to exist in

a variety of environments, including pathways associated with anaerobic metabolism. This is important because, although the bacteria require oxygen for growth in the laboratory, prolonged anaerobic survival might be necessary for long-term persistence in tissues⁴. The authors have also identified a series of genes that encode proteins involved in transcriptional regulation, indicating a potentially flexible response to the changing fortunes encountered during infection.

Koch was the first to comment on the unusual cell wall of *M. tuberculosis*, and its importance in mycobacterial physiology is reflected in a vast repertoire of genes involved in lipid and polysaccharide metabolism. The cell wall has been analysed biochemically by generations of outstanding researchers⁵, and there is real poetry in seeing how the complex structures of the components that make up the cell wall are mirrored in the myriad of genes that encode the corresponding biosynthetic machinery. Information about the genes involved in synthesis of the cell wall provides a cornucopia of potential drug targets.

The ability of *M. tuberculosis* to cause disease has not been ascribed to any well-defined virulence factors. However, Cole *et al.* have identified a repeated DNA sequence that encompasses a putative cell-entry protein, hinting at a parallel with pathogenicity islands found in other bacterial pathogens⁶. The genetic information required to cause

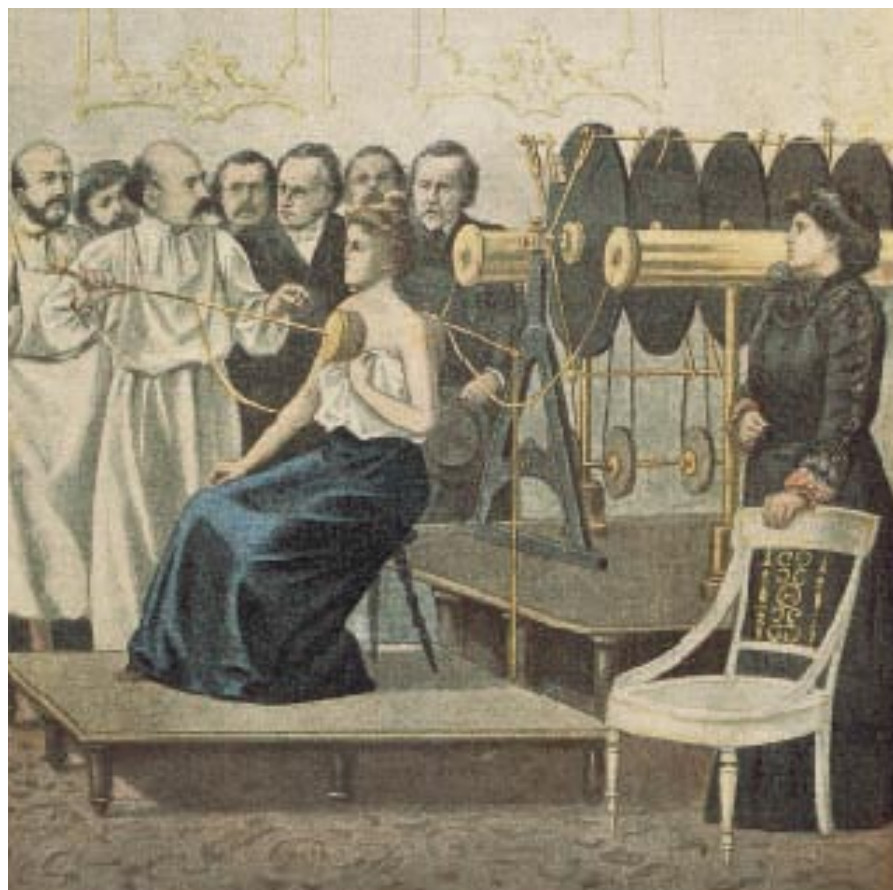


Figure 1 Attempt at treating tuberculosis by Francisque Crotte, 1901.

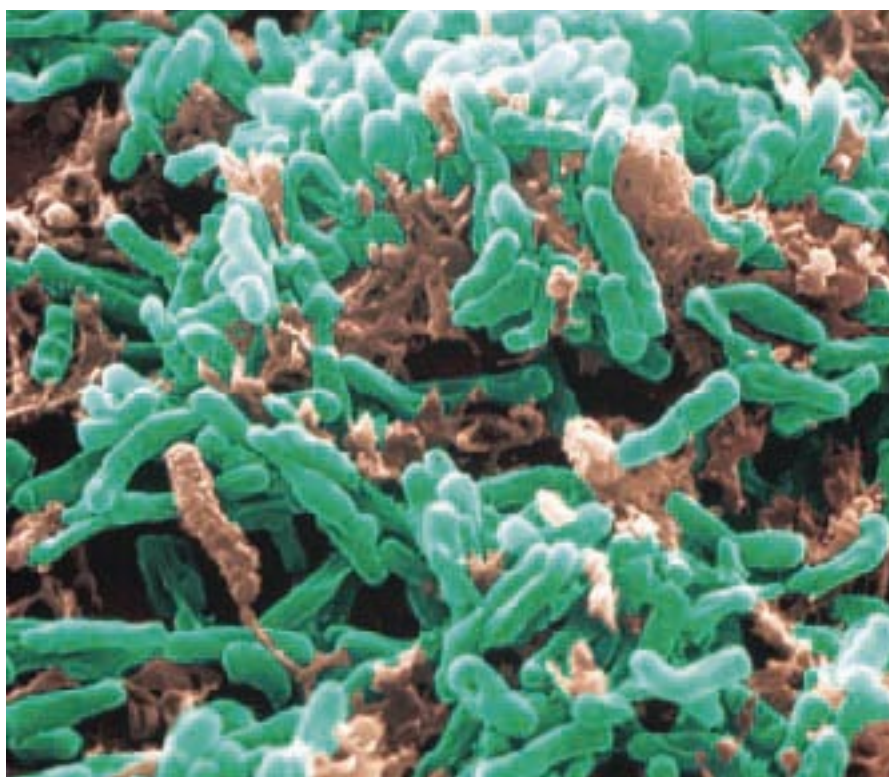


Figure 2 Electron micrograph of *Mycobacterium tuberculosis*, the genome of which has now been sequenced by Cole *et al.*². Also known as the white plague, tuberculosis was given the title “captain of all the men of death” by John Bunyan, towards the end of the seventeenth century.

disease is often exchanged between bacteria by a process termed ‘horizontal transfer’, and there is evidence that such transfer can occur in *M. tuberculosis*. The genome contains copies of mobile genetic elements known as insertion sequences, as well as genes derived from viruses that can attack bacteria, although there is no indication of a role in virulence. By comparing different strains of *M. tuberculosis*, the authors suggest that it probably acquired these genes before becoming established in its relatively isolated niche in the lung.

There is almost no diversity between one set of genes from a range of clinical isolates³, indicating a remarkably high level of sequence conservation in the *M. tuberculosis* genome. But DNA-fingerprinting techniques reveal that other parts of the genome are highly variable^{7,8}. An unexpected and exciting finding by Cole and colleagues is that one set of variable elements — the so-called polymorphic G+C-rich sequences — corresponds to a family of sequences that encode proteins with short peptide motifs made up of common repetitive domains. These proteins, which account for an astonishing 10% of the genome, are reminiscent of those implicated in antigenic variation in other microbes. By altering the pattern in which these proteins are expressed, pathogens present the immune system with a moving target, thereby increasing their chance of survival. Could the polymorphic G+C-rich sequences play an analogous role

for *M. tuberculosis*? Because the pathology of tuberculosis infection is probably mediated mainly by the misguided actions of the host immune response, rather than by bacterial toxins, the concept of some form of antigenic ‘smokescreen’ is intriguing.

The *M. tuberculosis* sequence is the first product of the Wellcome Trust Pathogen Genome Unit at the Sanger Centre, Cambridge, UK. The sequence of its cousin,

Mycobacterium leprae, is close to completion, and scientists at the Institute for Genome Research (TIGR) are working on other mycobacterial genomes, including a second isolate of *M. tuberculosis*. By comparing these genomes we should be able to identify genes associated with particular biological properties. Moreover, the recent development of tools for genetically manipulating *M. tuberculosis*^{9,10} will be invaluable in translating these bioinformatic ideas into the ‘wet lab’ products of new drug targets and attenuated strains for vaccine testing. Promising results with DNA vaccination against tuberculosis open up additional strategies in the search for protective antigens by whole-genome screens^{11,12}. The sequence data are being matched with maps of protein expression by two-dimensional gel electrophoresis and mass spectrometry, and with transcriptional analysis using microchip technologies. After several decades in the slow lane of classical microbiology, *M. tuberculosis* is once again at the cutting edge of science. □

Douglas B. Young is in the Department of Infectious Diseases and Microbiology, Imperial College School of Medicine, Norfolk Place, London W2 1PG, UK. e-mail: d.young@ic.ac.uk

1. Koch, R. *Berliner Klin. Wochenschr.* **19**, 221–230 (1882).
2. Cole, S. T. *et al. Nature* **393**, 537–544 (1998).
3. Sreevatsan, S. *et al. Proc. Natl Acad. Sci. USA* **94**, 9869–9874 (1997).
4. Wayne, L. G. & Hayes, L. G. *Infect. Immun.* **64**, 2062–2069 (1996).
5. Brennan, P. J. & Nikaido, H. *Annu. Rev. Biochem.* **64**, 29–63 (1995).
6. Covacci, A. *et al. Drug Dev. Res.* **41**, 180–192 (1997).
7. van Embden, J. D. *et al. J. Clin. Microbiol.* **31**, 406–409 (1993).
8. Poulet, S. & Cole, S. T. *Arch. Microbiol.* **163**, 87–95 (1995).
9. Bardarov, S. *et al. Proc. Natl Acad. Sci. USA* **94**, 10961–10966 (1997).
10. Pelicic, V. *et al. Proc. Natl Acad. Sci. USA* **94**, 10955–10960 (1997).
11. Huygen, K. *et al. Nature Med.* **2**, 893–898 (1996).
12. Tascon, R. E. *et al. Nature Med.* **2**, 888–892 (1996).

Electronic devices

Single electrons in silicon drops

D. Christian Glattli

More than 80 years after Millikan’s oil-drop experiment¹ establishing the quantization of the electric charge, there is an intense renewal of interest in charge-quantization effects at room temperature — but this time in tiny sub-micrometre conductors rather than in oil droplets. Why? At the current rate of miniaturization, metal-oxide field-effect transistors (MOSFETs) — the basic unit of modern computers — will run against fundamental limitations around the year 2015. One way around this limit might be found in a new type of field-effect transistor, called a single-electron transistor (SET), switchable by a single charge and operating under standard conditions (SETs that operate under sub-Kelvin temperatures already exist^{2,3}). Hence

the interest in the silicon-based device made by Lei Zhuang, Lingjie Guo and Stephen Chou⁴ (Fig. 1), which achieves SET performance at room temperature — and the fabrication process relies only on technology already developed for MOSFETs.

The conventional field-effect transistor is a narrow conducting channel connected to two low-resistance contacts, the source and the drain. The conductance of the channel can be varied efficiently by altering the voltage on a third electrode, the gate, which electrostatically attracts or repels charge carriers from the channel. In this way electrical signals can be amplified, and binary logic operations performed by turning the transistor on or off.

In the most popular FET, the MOSFET,

the channel is a nearly two-dimensional sheet of electrons (or holes) confined at the surface of a lightly p- (or n-) doped silicon substrate by the positive (or negative) voltage on a nearby gate. The gate is isolated from the channel by a thin silicon-oxide layer. The oxide's properties — easy growth, robustness and excellent electrical insulation — and the ability to put both n- and p-type channels on the same substrate are the reasons behind 30 years of successful miniaturization and large-scale integration⁵ of MOSFET technology.

But a practical limit to the miniaturization of the silicon transistor is expected within 10 to 15 years, when short-channel effects, leakage through a thin oxide layer, and dopant fluctuations become problematic. More fundamentally, when the device size goes below about 30 nm it will become comparable to the quantum-mechanical wavelength of the carriers, and quantum effects will dominate. Small size also implies few carriers, and maybe non-trivial charge-quantization effects. Nobody knows yet whether these new features can be exploited to improve the old FET, perhaps giving it new abilities, or whether they will simply degrade its performance.

But quantum-mechanical and charge-quantization effects are now commonly studied in metals and semiconductors at low temperatures. If these quantum devices can be pushed to operate at room temperature, they may be the answer to the miniaturization problem. One of the most intriguing devices is the SET, similar to the FET except that the channel is a small conducting island sandwiched between potential barriers. The barriers are high enough that charge carriers can only reach the island by tunnelling, a quantum-mechanical effect that allows particles to penetrate classically forbidden energy barriers. The operating principle is simple: to establish a current between source and drain, carriers must be added to or removed from the island. The sudden charge im-

balance implies an energy cost — the electrostatic energy of charging the island. If this charging energy is larger than the thermal energy, the SET is insulating. But by applying a voltage on the nearby gate, one can lower the potential energy of the island enough to reduce the charging energy below the thermal; the device becomes conducting.

Switching the current on and off with a gate voltage is nothing new, but there is more to the SET than that: going from a conducting to an insulating state corresponds to adding on average half a quantum of charge to the island. As a consequence, the conductance shows periodic oscillations with gate voltage. This characteristic has already been exploited to detect charges as small as 10^{-5} electrons, and SETs will soon be the most accurate current-frequency converters⁶.

The device of Zhuang and colleagues⁴ (Fig. 1) looks at first glance like a MOSFET. The channel between source and drain is an ultra-narrow silicon wire, patterned using electron-beam lithography. After subsequent growth of silicon oxide, the channel width is about 16 nm. Random fluctuations in the lithographic process lead to constrictions in the channel, which act as tunnel barriers, and sometimes two constrictions will isolate an island of about 12 nm diameter. That is used as the basis for the device.

The measured conductance shows the monotonic increase with gate voltage expected of a MOSFET, but, superimposed, there is a periodic variation with voltage characteristic of a SET. SET behaviour is clearly observed at room temperature, and fully develops at low temperature, where quantum confinement effects are also seen. The charging energy corresponds to about

1,200 K, as expected from the silicon island size (the tunnelling barrier is inversely proportional to island size). Previous semiconductor SETs barely reached a 200-K operating temperature.

This new result, combined with the recent demonstration of single-electron memory effects at room temperature, also in silicon^{7,8}, shows that a critical point has been reached in this field. Room-temperature single-electron tunnelling in purely metallic systems has also been reported⁹. Will the SET succeed the FET? That is not certain, as there are fundamental differences in their mode of operation. The FET gain is high, whereas that of the SET hardly reaches unity. And new molecular systems with conventional FET operation at room temperature, such as the recently demonstrated carbon-nanotube FET, are another promising route to miniaturization¹⁰. But perhaps some mixture of FET and SET technology will allow silicon to keep leading the way. □

D. Christian Glatelli is in the Service de Physique de l'Etat Condensé, CEA Saclay, F-91191 Gif-sur-Yvette, France.
e-mail: cglatelli@spec.saclay.cea.fr

1. Millikan, R. A. *Phys. Rev.* **32**, 349 (1911).
2. Fulton, T. A. & Dolan, G. J. *Phys. Rev. Lett.* **59**, 109–112 (1987).
3. Grabert, H. & Devoret, M. H. (eds) *Single Charge Tunneling* (Plenum, New York, 1992).
4. Zhuang, L., Guo, L. & Chou, S. Y. *Appl. Phys. Lett.* **72**, 1205–1207 (1998).
5. Fowler, A. *Phys. Today* October, 50–54 (1997).
6. Keller, M. W., Martinis, J. M., Zimmerman, N. M. & Steinbach, A. H. *Appl. Phys. Lett.* **69**, 1804–1806 (1990).
7. Guo, L., Leobandung, E. & Chou, S. Y. *Science* **275**, 649–652 (1997).
8. Tiwari, S. et al. *Appl. Phys. Lett.* **68**, 1377–1379 (1996).
9. Shirakashi, J. & Matsumoto, K. *Appl. Phys. Lett.* **72**, 1893–1895 (1998).
10. Tans, S. J., Verschuereen, A. R. M. & Decker, C. *Nature* **393**, 49–52 (1998).

Evolutionary biology

Help and you shall be helped

Régis Ferrière

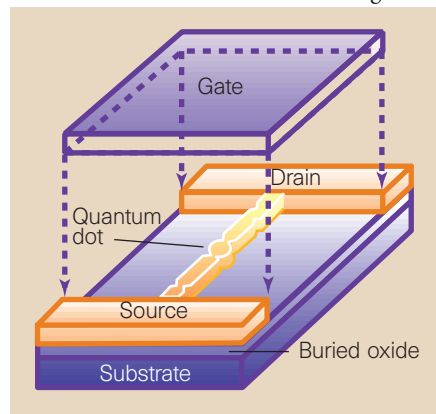


Figure 1 A new single-electron transistor (SET). At its core is a 'quantum dot', an island about 12 nm across, which is inaccessible to electrons unless a voltage is applied in the gate, above. This is the first SET to work at room temperature.

How do moral systems evolve? The common view, rooted in game theory (see box, overleaf), is that cooperation and mutual aid require tight partnerships among individuals, or close kinship¹. But this dogma is now being shaken and, on page 573 of this issue, Nowak and Sigmund² report a new mathematical model to show that cooperation can become established even if recipients have no chance to return the help to their helper. This is because helping improves reputation, which in turn makes one more likely to be helped.

Nowak and Sigmund's stance differs radically from classical approaches, because they assume that any two organisms are unlikely to interact twice. This excludes any direct reciprocity, meaning that you should not expect your assistance to be returned by somebody you helped on a previous

encounter, but by somebody else. For such indirect reciprocity to work, individuals must assess others in their group by watching their games from the sidelines and assigning 'image scores' to them. If you watch one individual helping another, you can allot him or her one point of image score; but if you witness the individual refusing to help, withdraw one point. Then, when you are asked to help, you will cooperate if your partner's score is high — otherwise, you defect (that is, you don't cooperate). Strategies can be more or less discriminating according to their 'score cut'. For example, a 'cooperative' strategy has a score cut of zero or less, meaning that the individual then cooperates with other individuals on their first interaction.

This yields a new dilemma. When cooperating, your only reward is a score rise, at the cost of helping. You bet that the benefit of

helping will be returned by someone who acknowledges your good reputation. This works, like an insurance principle, as soon as enough clients subscribe to it. On the other hand, if you punish a low scorer by defecting, this blemishes your own reputation. The punishment is possibly unfair — if the partner is a discriminator who had the bad luck to encounter low-score players, you will not help him. (This is in contrast to strategies based on reputation, where a defection provoked by the co-player's misbehaviour inflicts no black mark³.)

To probe the success of discriminating strategies, Nowak and Sigmund² have simulated the effect of natural selection on a population that includes various levels of discrimination. These range from blind cooperation (where every co-player receives help, no matter what his score) to unconditional defection (where no co-player is ever considered worthy of help). In every generation, there is a given — usually low — number of pair-wise interactions, and the frequency of a strategy increases in proportion to its pay-off. From their simulations, the authors show that the winner is often the cooperative strategy that discriminates the most, the one whose score cut equals zero. That way, cooperation is established even though repeated encounters almost never occur.

The dynamics of the game become more complicated if reproduction is not quite faithful (there may be mutations in the offspring's score cut). Then, the regime of cooperation is occasionally punctuated by

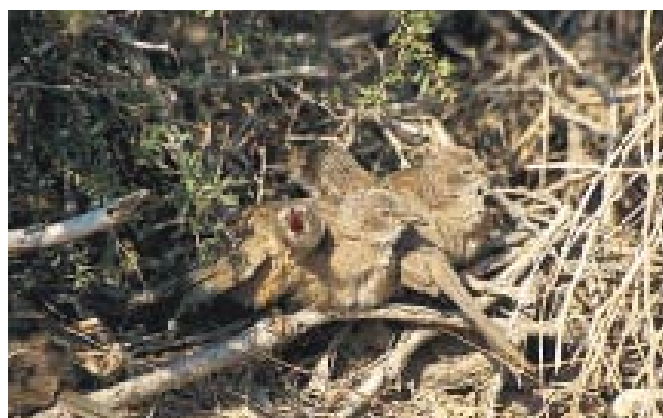


Figure 1 Babblers jostling to give help — an instance of indirect reciprocity?

bursts of defection, and the underlying mechanism is always the same. Unconditional cooperators spread by drift, and then defectors who can exploit them invade. In most cases, the discriminators return and re-establish cooperation. Surprisingly enough, a simplified model shows that the only way defectors can achieve a lasting triumph is by invading very rarely, because this provides time for stochastic fluctuations to eliminate almost all discriminators. A population must be challenged often if it is to remain immune to defection.

Individuals may not be able to keep track of all scores. Discriminators will spread provided the amount of information that they have about co-players exceeds some threshold. In Nowak and Sigmund's model, this amount is externally imposed. But it should reflect individual features such as the players' cognitive abilities, their propensity to move

and the cost incurred for monitoring others. Spatial models could assess the effect of such traits⁴. For example, reduced mobility might be detrimental, unless it is compensated for by an exchange of information, such as through gossip⁵ (the speed of rumours is proverbial).

The discriminator strategy has an antecedent — the 'observer tit for tat' (OTFT), developed in the context of an iterated prisoner's dilemma⁶. The OTFT strategy prescribes defection in the first round if the opponent was seen defecting in his last interaction, and thereafter the standard tit for tat. OTFT can hold its own against defectors when the likelihood of repeated interactions decreases. This suggests that a gradual transition from direct to indirect reciprocity may occur as partnerships become looser.

Richard Alexander pioneered the concept of indirect reciprocity to deal with moral systems in human societies⁷. But what about other species? The Arabian babbler, *Turdoides squamiceps* (Fig. 1), a gregarious kind of bird, seems to enjoy helping other babblers⁸. More than that, the babblers actually compete for the position of the donor. Obviously the increase in status compensates for the cost of cooperating. Also, a discriminating choice of partner is sometimes involved in the associations of plants and bird pollinators⁹. There might not be much morality in birds or plants, yet indirect reciprocity based on image scoring is a tempting explanation for some of their mutualistic inclinations. □

Régis Ferrière is in the Laboratoire d'Ecologie, CNRS UMR 7625, Ecole Normale Supérieure, 46 rue d'Ulm, 75230 Paris cedex 05, France.
e-mail: rferrier@hall.snv.jussieu.fr

1. Axelrod, R. & Hamilton, W. D. *Science* **211**, 1390–1396 (1981).
2. Nowak, M. A. & Sigmund, K. *Nature* **393**, 573–577 (1998).
3. Sugden, R. *The Evolution of Rights, Co-operation and Welfare* (Blackwell, Oxford, 1986).
4. Ferrière, R. & Michod, R. E. *Am. Nat.* **147**, 692–717 (1996).
5. Enquist, M. & Leimar, O. *Anim. Behav.* **45**, 747–757 (1993).
6. Pollock, G. B. & Dugatkin, L. A. *J. Theor. Biol.* **159**, 25–37 (1992).
7. Alexander, R. D. *The Biology of Moral Systems* (Aldine de Gruyter, New York, 1987).
8. Zahavi, A. *J. Avian Biol.* **26**, 1–3 (1995).
9. Bull, J. J. & Rice, W. R. *J. Theor. Biol.* **149**, 63–74 (1991).
10. Sigmund, K. *Games of Life* (Oxford Univ. Press, 1993).

Game theory and the prisoner's dilemma

The mathematical theory of games (evolved from poker and chess and intended in the first place to model social and military conflicts) explains all sorts of animal behaviour: the restrained nature of fighting between stags, for instance, or the battle of the sexes between female and male guppies¹⁰.

The main concepts of game theory are pay-off and strategy. A strategy is a programme for the player — a sequence of decisions. When the game is over, each player receives a pay-off which depends on the player's decisions as well as those made by opponents. In biological games, pay-off simply

means an increase in fitness (the number of offspring, for example). To model the effect of natural selection on a population that includes several strategies, the game is played in every generation. From one generation to the next, the frequency of a strategy increases or decreases in proportion to its pay-off.

Game-theoretical models are typified by the well-trodden prisoner's dilemma game^{1,10}. This problem originally considers two suspects, imprisoned separately, each contemplating whether to keep silent (to cooperate) or to sing (to defect). Pay-offs are measured in length of confinement. If

both prisoners cooperate, they each do better (lighter sentence) than if they both defect. Yet each could do better by defecting than by cooperating. Both parties expect an advantage from the exchange; but this advantage would be even greater if one's own contribution were omitted, or at least a bit reduced. If the game is played only once, defection is surely the best strategy. But if the same opponents play repeatedly, this allows each player to return the co-player's help or punish his or her defection¹. Thus, the expectation of further dealings can make cooperative strategies more advantageous. **R.F.**



100 YEARS AGO

... owing to the kindness of Dr. Hampson, we have been furnished with about 750 cubic centimetres of liquid air, and, on allowing all but 10 cubic centimetres to evaporate away slowly, and collecting the gas from that small residue in a gas-holder, we obtained, after removal of oxygen with metallic copper and nitrogen with a mixture of pure lime and magnesium dust, followed by exposure to electric sparks in presence of oxygen and caustic soda, 26.2 cubic centimetres of a gas, showing the argon spectrum feebly, and, in addition, a spectrum which has, we believe, not been seen before. ... it may be concluded that the atmosphere contains a hitherto undiscovered gas with a characteristic spectrum, heavier than argon, and less volatile than nitrogen, oxygen, and argon; the ratio of its specific heats would lead to the inference that it is monatomic, and therefore an element. If this conclusion turns out to be well substantiated, we propose to call it "krypton", or "concealed." Its symbol would then be Kr.

From *Nature* 9 June 1898.

50 YEARS AGO

The practice of the whalers is to try to catch the largest whales – the gunner who does this gains in reputation; but any whale is better than none, since the factory ships have to be filled with oil. Since the average length of the whales killed is decreasing, it must be concluded that either the larger whales are becoming more difficult to catch or that they have been pretty well killed off. The former assumption cannot stand, for against powerful modern catchers the bolting whale stands no chance. It must take frequent breaths while travelling at speed, remaining practically at the surface all the time, and so can be easily followed and run down. With regard to the latter assumption, the remarks made to me after the last Antarctic season by a very experienced whaler are pertinent. He told me that blue whales were not what they used to be, and that the enormous animals of past years were either extremely scarce or non-existent. ... The inevitable conclusion is that the larger animals have been killed off, and that the survivors do not live long enough to attain the great size of former years.

From *Nature* 12 June 1948.

Observational cosmology

Through a glass brightly

Andrew Blain

The record for the most distant object in the Universe is broken regularly, but the record for the object with the largest apparent luminosity has been much more durable, resting¹ with IRASF10214+4724 since 1991. But a new champion has emerged — as described by Irwin *et al.*² in a forthcoming paper, APM08279+5255 appears to be ten times more luminous.

F10214 was identified during a survey to measure the redshifts of galaxies detected using the IRAS satellite. IRAS was sensitive to the far-infrared thermal emission from grains of interstellar dust in the Milky Way and other galaxies. The dust is heated to between 20 and 100 K by the visible and ultraviolet light emitted by stars or active galactic nuclei (AGN), and re-radiates this energy in the far-infrared waveband. So the presence of dust decreases the visible and ultraviolet luminosity of a galaxy or AGN, and increases its infrared luminosity.

The most luminous IRAS galaxies mostly have redshifts of up to about 0.1, but F10214 was found to have the much larger redshift of 2.3. Sources with large redshifts are at great distances — the light observed from objects at redshifts 0.1 and 2.3 was emitted when the Universe was 85% and 17% of its present age, respectively. To be detectable at such a large distance, F10214 had to be about 100 times more luminous than its low-redshift counterparts, and so appeared to have a luminosity of approximately $3 \times 10^{14} L_{\odot}$ (a solar luminosity, $L_{\odot}$, is 3.9×10^{26} W). Taken at face value, this would be a very remarkable object. The excitement was tempered slightly when F10214 was found to be magnified by a factor of about 40 by the gravitational lens-

ing³ effect of a foreground galaxy⁴. Its true luminosity is only comparable to that of the most powerful low-redshift IRAS galaxies. Nevertheless, it provided a valuable case study as the first distant galaxy found to emit most of its energy in the far-infrared waveband, fitting in well with models of galaxy formation which predict large quantities of dust in early objects.

If F10214 surprised its observers, then APM08279 must have been a rude shock for Irwin *et al.*². It was discovered during spectrographic observations made to measure the exact velocities of cool carbon-rich stars orbiting in the halo of the Milky Way⁵. APM08279, first thought to be a star, was recognized as a distant quasar, at a redshift of 3.87, because of the characteristic redshifted emission and absorption lines in its spectrum.

As bright as a star in the halo of our Galaxy, but at the outer edge of the observable Universe, APM08279 appears to be remarkably luminous. Its relative brightness in the APM and IRAS catalogues indicates that more than half of its power is emitted in the mid-infrared waveband, and its inferred luminosity is about $5 \times 10^{15} L_{\odot}$. On average, one IRAS source with the same apparent magnitude was detected per two square degrees of the sky — the area of about ten full Moons. No more than one in 10^4 , and maybe as few as one in 10^6 of these sources, is expected to have a redshift greater than three⁶, and so fewer than two sources like APM08279 would be expected in the whole sky.

But APM08279 is almost certainly gravitationally lensed by a foreground galaxy. Absorption lines in its spectrum show that our line of sight to the quasar intersects three

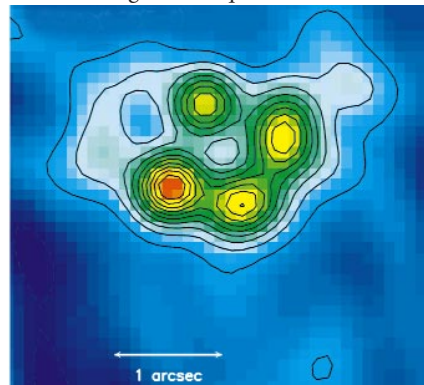
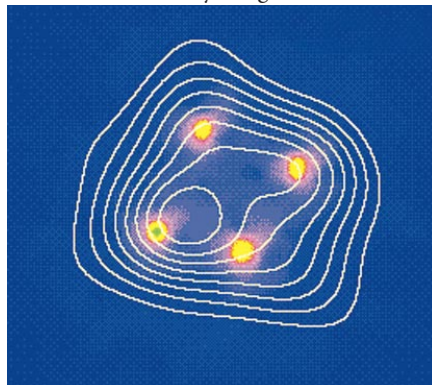


Figure 1 The Cloverleaf. This quasar has been gravitationally lensed, by an intervening galaxy, into four separate images. As with the newly discovered ultra-luminous object APM08279+5255, the gravitational lensing makes the Cloverleaf seem much brighter than it really is. The optical emission (left) from the centre of the object heats surrounding dust, which re-radiates the energy in the sub-millimetre band (right).

IRAM

potential lensing galaxies. Furthermore, its optical image appears to be slightly elongated². This can be explained if the source is a pair of almost equally bright lensed images with a separation of 0.33 arcsec, blurred into one by the atmosphere — exactly as expected for a strongly magnified quasar. When a correction is applied for the estimated lensing magnification, the source becomes fainter, and so less unusual.

The higher 0.1-arcsec spatial resolution of the Hubble Space Telescope should confirm or refute the lensing hypothesis. Another quasar, H1413+117 — ‘the Cloverleaf’, is lensed by an elliptical galaxy into four images of comparable brightness⁷ (Fig. 1). The Cloverleaf is also a bright IRAS source, and has been observed in detail over a wide range of wavelengths. It is to be hoped that observations of comparable quality for APM08279, which is brighter than the Cloverleaf, will be available soon.

The identification of APM08279 is more strong evidence that luminous dusty galaxies and quasars exist in the distant Universe. Many more such sources should still be lurking unconfirmed in the IRAS catalogue². In fact, it is quite plausible that IRAS sources as dusty and luminous as APM08279, but with more heavily obscured spectra similar to that of F10214, lie in the surveyed field², but were not observed because they are missing from the optically selected APM catalogue. Accounting for this class of source would increase the numbers of high-redshift, high-luminosity galaxies even further.

The existence of sources such as APM08279 is consistent with the results of the first direct systematic survey that is

sensitive to very distant dusty galaxies and quasars⁸. This survey was carried out in the submillimetre waveband, to detect the redshifted radiation from distant dusty star-forming galaxies and quasars. Submillimetre-selected sources are almost all located at high redshifts, at which gravitational lensing is most likely to occur. It is difficult to build the sensitive instruments required for submillimetre surveys⁹, but they are now becoming available¹⁰, and the planned all-sky survey by ESA's Planck Surveyor satellite will be of particular interest. Planck Surveyor will not only map the cosmic microwave background radiation with unprecedented precision, but will also detect a large number of extremely distant and luminous gravitationally lensed galaxies and quasars¹¹. The identification of APM08279 is an excellent omen for this systematic survey of the distant Universe. □

Andrew Blain is in the Cavendish Laboratory, Madingley Road, Cambridge CB3 0HE, UK.
e-mail: awb@mrao.cam.ac.uk

1. Rowan-Robinson, M. *et al.* *Nature* **351**, 719–721 (1991).
2. Irwin, M. J., Ibata, R. A., Lewis, G. F. & Totten, E. J. *Astrophys J. Lett.* (in the press).
3. Schneider, P., Ehlers, J. & Falco, E. E. *Gravitational Lenses* (Springer, Berlin, 1992).
4. Close, L. H., Hall, P. B., Liu, C. T. & Hege, E. K. *Astrophys J. Lett.* **452**, L9–L12 (1995).
5. Totten, E. J. & Irwin, M. J. *Mon. Not. R. Astron. Soc.* **294**, 1–27 (1998).
6. Blain, A. W., Smail, I., Ivison, R. J. & Kneib, J.-P. *Mon. Not. R. Astron. Soc.* (submitted); astro-ph/9806062 on xxx.lanl.gov
7. Kneib, J.-P. *et al.* *Astron. Astrophys.* **329**, 827–839 (1998).
8. Smail, I., Ivison, R. J. & Blain, A. W. *Astrophys J. Lett.* **490**, L5–L9 (1997).
9. Philipps, T. G. (ed.) *Advanced Technology MMW, Radio and Terahertz Telescopes. Proc. SPIE 3357* (in the press).
10. Holland, W. S. *et al.* in *Advanced Technology MMW, Radio and Terahertz Telescopes. Proc. SPIE 3357* (ed. Philipps, T. G.) (in the press).
11. Blain, A. W. *Mon. Not. R. Astron. Soc.* (in the press); astro-ph/9801098 on xxx.lanl.gov.

Thermodynamics

Liquid landscape

Austen Angell

Supercooled liquids — that is, liquids at temperatures below their normal freezing point — can undergo a subtle transition to a microscopically fixed, yet amorphous state: a glass. The temperature of this transition depends on how quickly the liquid is cooled, so it has seemed more natural to describe the process in terms of kinetics, rather than some immutable thermodynamics. But on page 554 of this issue¹ Sastry, Debenedetti and Stillinger show how changes in the dynamic properties of glass-forming liquids^{2,3} can be related to changes in the nature of the system's underlying energy landscape.

‘Landscape paradigms’ are used to describe the qualitative behaviour of complex systems⁴. In such systems there are many metastable states, and many possible transitions between the different states. The system's state is represented by a point which

moves on or above a surface, according to rules decided by the problem under consideration. The types of system described by the landscape model range from economics and evolution, through neural networks, spin glasses and proteins, to molecular clusters and viscous liquids. For each problem, the height of the surface may correspond to a different parameter — ‘fitness’ is to be maximized for an evolving species, for example; and some energy is to be minimized for a system of molecules. For the latter, the landscape topology controls the kinetics⁵, and the average energy of the minima visited at temperature T reflects the thermodynamics.

For viscous liquids and glasses, the system point depends on how a collection of N particles are arranged, because the system's energy is determined by the exact positions of all the particles. The energy landscape is a surface with $3N+1$ dimensions,

and impossible to conceptualize properly. Nevertheless, a two-dimensional representation (Fig. 1a) is useful, and can illustrate the distinctions between liquid, crystal and glass.

In equilibrium, the liquid moves between minima in a region of energy determined by the temperature. When the temperature is increased, the system will spend most of its time ‘higher up’ on the energy landscape, gaining access to more possible states. This is in accordance with the principle that fixed-volume systems in equilibrium should minimize their free energy (the Helmholtz free energy $A = E - TS$, where E is the internal energy and S is the entropy). S is large when the number of states that the system can visit is large. The relation is that inscribed on Boltzmann's tomb, $S = k_B \ln W$ where W is the number of possible states (here, energy minima on the landscape; vibrational states equilibrate far more quickly, and so can be treated separately).

Perfect crystals occupy just one deep minimum, and melt in order to lower their free energy by gaining access to all the minima of the landscape — for melting at fixed volume, the entropy change TdS just balances the internal energy change dE . Glass-formers are liquids that fail to find their way back to the crystal minimum on cooling, and hence wander down among the myriad minima of the landscape as they supercool (‘down’

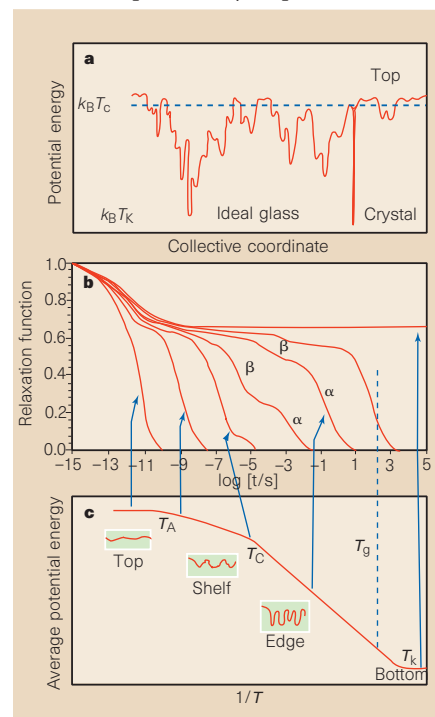


Figure 1 Three angles on glass. a, In the energy-landscape depiction, the internal energy of a system depends on the configuration of all its particles (here simplified to one dimension). b, The relaxation function shows the approach to equilibrium over time. c, The average energy of the energy-landscape slice sampled at different temperatures, with an indication (in boxes) of the form of the minima in each energy range.

because there are then fewer minima accessible, and S is smaller). Eventually the system gets stuck in one of the lower minima — not as low as for the crystal — and we have a glass.

The system gets stuck because the vibrational energy (which adds on to the landscape energy to 'float' the system a little above the bottom of any minimum it occupies) becomes insufficient to activate the system from one minimum to the next. Bringing the temperature up a little allows the system to slowly explore the surrounding minima, hence to restore an equilibrium state (minimum in A). This is called relaxation, annealing or ageing depending on whether one works on liquids, glasses or polymers.

To ferret out the details of this picture for a model potential, Sastry *et al.*¹ have used a system containing a few hundred argon atoms of two different sizes, which interact in a way that stabilizes the liquid and makes it impossible to find any crystal minimum^{3,5,6}. Sastry *et al.* equilibrate the system at various temperatures, keeping the volume constant. Then, to determine the internal energy of the 'inherent structure' (the structure sampled most frequently by the system at a given temperature), they quench the system, removing the vibrational energy by a technique that ensures a direct descent to the bottom of the energy minimum above which the system was oscillating at the moment chosen. If all minima have the same depth, then the final energy should be independent of temperature. Indeed, that is what is found, within close limits, for high temperatures — such as those above normal melting points. This

means that there can't be a statistically interesting number of higher-energy states anywhere on the landscape, or else the TS factor would drive the system to sample them.

But for supercooled liquids below a well-defined temperature (T_A in Fig. 1), the energy of the inherent structure starts to decrease as the temperature of equilibration is lowered. T_A turns out to be just the temperature at which earlier simulations³ found a departure from exponential relaxation (and the development of a second stage in the total relaxation process; Fig. 1) and the onset of a special temperature-dependence considered characteristic of viscous liquids. This regime is described well by the much-studied and highly mathematical 'mode-coupling theory'⁷, which predicts a power-law variation of relaxation time with T . The physical picture is one of particles 'caged' by their neighbours but eventually escaping on the relaxation timescale.

More interesting is the landscape 'slice' visited as the temperature decreases further, to where mode-coupling theory starts to break down. The breakdown is so far not seen in the simulations because of computing speed limitations, but is known from laboratory studies. Sastry *et al.*¹ found that, although their system also got stuck in this range, the landscape where it stuck was characterized by suddenly deeper minima, and, moreover, minima of a simpler form (Fig. 1c).

A break in the 'average energy sampled' versus T , such as occurs near T_c in Fig. 1c, must mean a break in the number density of

the energy minima, as it is entropy that drives the system upward in energy as T increases. We use a $1/T$ scale to simplify the drawing of connections between the slice of landscape frequented by the system at temperature T and the character of the relaxation process.

Figure 1 shows how the energy landscape discussed by Sastry *et al.* relates to the relaxation function^{2,3}. Short-timescale relaxation data^{2,3}, and energy 'slice' data from the present study, are joined with the longer-timescale findings of experimentalists at lower temperatures down to and below the glass transition. Part c shows how the average sampled energy bottoms out at a temperature characteristic of the energy of the lowest minimum on the landscape of part a. This is the Kauzmann temperature (T_K), where the non-vibrational entropy vanishes for infinitely slow cooling — the 'ideal glass temperature' of theory^{8,9}. Figure 1b shows how two-step relaxation begins at about the same temperature as a weakly sloping energy profile begins. Arrival at the end of this 'continental shelf' (at T_c) is accompanied not only by the onset of ruggedness (Fig. 1c), but by a new complexity in the relaxation function. This is the little-understood splitting of the de-caging relaxation into two parts (α and β) of very different T dependence, as predicted by Goldstein¹⁰. Is this, and the shelf edge, due to cage-cage percolation¹¹ to give a domain structure? The simulations are not yet fast enough to resolve this question.

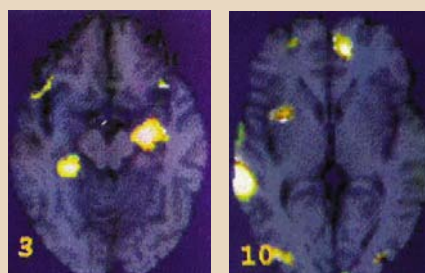
In the meantime, we need to understand liquid 'fragility' (sensitivity to temperature change) in landscape terms¹². Comparable

Neurobiology

Brain, heart and stress

Positron emission tomography (PET) brain scanning is being used to tackle ever more subtle scientific and medical questions. Witness the two images here, which come from work carried out by Robert Soufer and colleagues of Yale University and are reported in *Proceedings of the National Academy of Sciences* (95, 6454–6459; 1998). They respectively show hyperactivation (left) and deactivation of certain parts of the brain, as measured by blood flow, in subjects with coronary artery disease when under mental stress.

The rationale behind this work is that although stress is associated with heart attacks and other cardiac problems in people with coronary artery disease, research has hitherto centred on the vascular and psychological aspects. The brain is undoubtedly involved in some way — through stress-induced activation of certain regions leading, perhaps, to a response from hormones or the sympathetic nervous system that affects heart function. An end result can be



myocardial ischaemia, with the heart muscle being provided with inadequate oxygen to function properly.

Soufer *et al.* took PET scans of patients with coronary artery disease, controlled against healthy people, when under mental stress and not under stress. They monitored heart function at the same time. The stress involved mental arithmetic, serial subtraction of numbers, the subjects being prompted for ever faster performance and their mistakes being corrected in no uncertain terms.

The resulting differences in patterns of blood flow and therefore brain activation

or deactivation were highly complicated and are open to various interpretations. Most notably, however, as depicted in the images shown here, in three of the ten patients with coronary artery disease significant differences in activation/deactivation when under stress were associated with indications of myocardial ischaemia. For instance, the image on the left here shows hyperactivation in the left hippocampus and right parahippocampal regions; that on the right, deactivation in the cingulate gyrus among other areas.

This work can only be considered as a pilot study, but it does show that particular parts of the brain mediate the effects of stress on heart function in people suffering from coronary artery disease. The implications for devising possible treatments of the disease with drugs or changes in behaviour or lifestyle are clear. Realizing those prospects is a distant goal, but a start has been made.

Tim Lincoln

simulations should show whether the sparser minima of non-fragile systems 'eat up' the shelf seen in the fragile-liquid sampled-energy profile (Fig. 1c), or simply push the Kauzmann temperature to lower values. Fragility is minimized in open networks, so can probably be 'tuned down' by adding variable-strength tetrahedral potentials (as in tin and silicon) to the simple atomic potentials of the present study. That should increase liquid 'strength' and yet maintain low crystallization probability. Such studies may provide the key to filling in the rest of the picture for systems with covalent, metallic or ionic interactions. □

Austen Angell is in the Department of Chemistry and Biochemistry, Arizona State University,

Box 87 1604, Tempe, Arizona 85287-1604, USA.
e-mail: caa@asu.edu

1. Sastry, S., Debenedetti, P. G. & Stillinger, F. H. *Nature* **393**, 554–557 (1998).
2. Colmenero, J. *et al.* *Phys. Rev. Lett.* **78**, 1928–1931 (1997).
3. Kob, W. & Andersen, H. C. *Phys. Rev. E* **51**, 4626 (1995).
4. Frauenfelder, H. *et al.* (eds) *Landscape Paradigms in Physics and Biology* *Physica D* spec. issue **107** (1997).
5. Donati, A. *et al.* *Phys. Rev. Lett.* **80**, 2338–2340 (1998).
6. Weber, T. A. & Stillinger, F. H. *Phys. Rev. B* **32**, 5402 (1985).
7. Götze, W. in *Liquids, Freezing, and the Glass Transition* (eds Hansen, J.-P. & Levesque, D.) 287–503 (NATO-ASI, North-Holland, Amsterdam, 1991).
8. Adam, G. & Gibbs, J. H. *J. Chem. Phys.* **43**, 139–146 (1965).
9. Parisi, G. preprint cond-mat/9712079 on xxx.lanl.gov.
10. Goldstein, M. *J. Chem. Phys.* **51**, 3728–3739 (1969).
11. Novikov, V. N., Rössler, E., Malinovsky, V. K. & Surovtsev, N. V. *Europhys. Lett.* **35**, 289 (1996).
12. Richert, R. & Angell, C. A. *J. Chem. Phys.* **108**, 9016–9027 (1998).

Physiology

Chemokines beyond inflammation

Richard Horuk

The old adage “don’t judge a book by its cover” is certainly apt when applied to chemokines. Although they were originally defined as host defence proteins, chemokines clearly have other functions — growth-regulatory and angiogenic properties, for example — that extend well beyond the regulation of leukocyte migration^{1,2}. Moreover, chemokine receptors such as CXCR4 are important¹ in the pathogenesis of the human immunodeficiency virus, HIV-1. A further level of complexity in chemokine biology is now revealed by Tachibana *et al.*³ and Zou *et al.*⁴ on pages 591 and 595 of this issue. They report that deletion of the *CXCR4* gene is embryologically lethal in mice, producing a multiplicity of effects including serious developmental defects in the immune, circulatory and central nervous systems. These studies expand the biological importance of chemokines, from that of simple immune modulators to a much broader biological role than was at first appreciated (Fig. 1).

The CXCR4 chemokine receptor is expressed throughout the body, and is a co-receptor⁵ for cellular entry by T-cell-tropic but not macrophage-tropic strains of HIV-1. Its ligand is the stromal derived factor SDF-1 (ref. 6), which is known to be important in embryonic development⁷. Mice that lack the *SDF-1* gene die *in utero*, with severe defects in the ventricular septum of the heart. They also have fewer B-lymphocyte progenitors in the liver and bone marrow, suggesting defects in the formation of red and white blood cells. An obvious question is whether deletion of the CXCR4 receptor would produce a similar profile of deficiencies, and this is now addressed by Tachibana *et al.* and Zou *et al.* Although they find that CXCR4-deficient mice have a similar phenotype to the *SDF-1* knockouts, the results also hold some surprises.

The two groups show that when both copies of the CXCR4 gene are knocked out, around half the mice die *in utero* by the 18th day of embryological development (E18.5), the remainder dying within one hour of birth. The similarity of this phenotype to that of the SDF-1-knockout mice confirms that CXCR4 and SDF-1 are a specific receptor–ligand pair that act in cardiac and blood-cell development. But the real surprise to come from the knockout mice is that CXCR4 is involved in the embryological development of neuronal networks in the central nervous system (CNS) and in the development of blood vessels in the gastrointestinal tract.

Zou *et al.*⁴ compared brains from normal and CXCR4-deficient mice, and found abnormalities in the architecture of the cerebellum in the knockout mice. Specifically, cells from the external granule layer prematurely migrated into the internal granule layer — a process that normally does not occur until after birth. The generation of precisely formed neural networks, also called neuronal patterning, is important for proper functioning of the entire CNS, ensuring cell-to-cell communication through a system of correctly formed cellular synapses. During normal development, neurons migrate from the germinal matrix to specific positions within the CNS. Cells in the external granule layer migrate from the surface of the cerebellum, through the molecular and Purkinje-cell layers, to form the internal granule layer. Scaffolding provided by radial glia helps these granular neurons in their migration. In the CXCR4-deficient mice, however, the authors found that this orderly process is severely disrupted by the premature migration and abnormal clustering of neurons, despite the presence of intact radial glia.

Thus, the work of Zou *et al.* greatly extends previous studies that have shown

neuronal expression of CXCR4. This receptor is clearly important in the development of the CNS, but what role could it play? There are several possibilities. First, there is compelling evidence for cross-talk between serpentine receptors⁸ (the superfamily to which CXCR4 belongs). Signals generated from one receptor can either stimulate or inhibit signalling by another. Thus, CXCR4 may regulate migration of cells from the external granule layer by inhibiting their ability to respond to other chemoattractants — in fact, neurons can migrate in response to chemokines⁹. Alternatively, SDF-1 has been shown¹⁰ to induce apoptosis in a human neuronal cell line through CXCR4, so perhaps CXCR4 aids in the apoptotic elimination of cells that have migrated incorrectly in the CNS. This would help to ensure correct neuronal patterning. To learn more, however, we need to identify the cell types in the CNS that express CXCR4 during development.

The formation of new blood vessels can be divided into three stages: vasculogenesis, which involves the maturation of mesodermal precursor cells into haemangioblasts; angiogenesis, in which these cells develop into an initial capillary network; and, finally, pruning and remodelling to form a functional circulatory network. Tachibana *et al.*³ show that part of this process is defective in the gastrointestinal tract of CXCR4-deficient mice. Development of the vascular system in the gastrointestinal tract of both normal and CXCR4-knockout embryos was identical initially, and a highly vascularized network was observed at E11.5. However, at E13.5 there were clear differences in the circulatory networks of the two groups. The normal embryos developed large and small branches of the superior mesenteric arteries and veins, whereas in the CXCR4-knockout animals the larger branches of these vessels were missing, even though the vessels themselves were normal. Although these deficiencies did not affect development of the gastrointestinal tract, the abnormal circulatory system in the knockout embryos led to haemorrhaging and intestinal obstruction. As expected, SDF-1-knockout animals had a similar phenotype.

The CXCR4 receptor seems to be involved in vascular development in an organ-specific way, because the vasculature of other organs — including the brain and heart — was normal. In contrast to tyrosine kinase receptors such as vascular endothelial-cell growth factor and TIE-1, which seem to be important throughout vascular development, the CXCR4 receptor has a more restricted role. It remodels and prunes vessels, leading to the formation of a correctly branched gastrointestinal circulatory network. The discovery that SDF-1 acts in vascular development is not totally surprising, because other chemokines from the

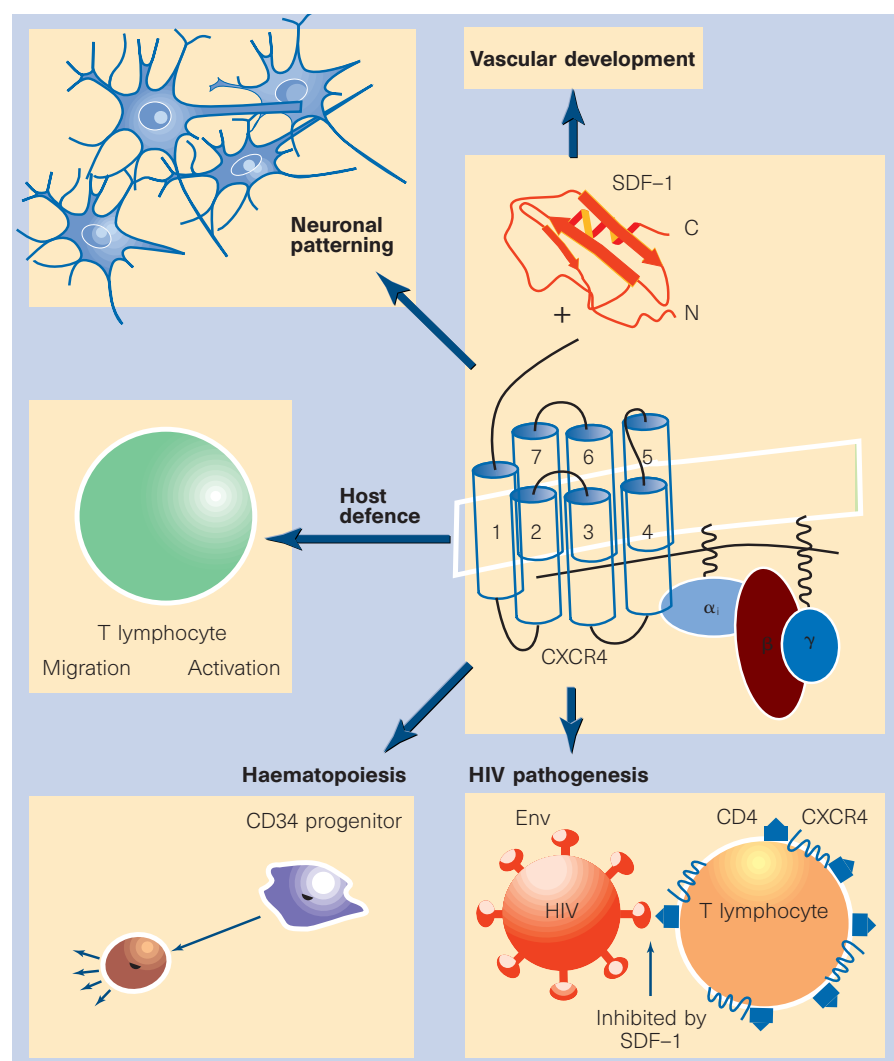


Figure 1 Biological roles of the chemokine receptor CXCR4 and its ligand SDF-1. Tachibana *et al.*³ and Zou *et al.*⁴ have found that mice lacking CXCR4 have a similar phenotype to those lacking SDF-1 — namely, serious developmental defects in the immune and circulatory systems. Surprisingly, the mice also show defects in the central nervous system and in the formation of blood vessels within the gastrointestinal tract, indicating that chemokines have much more widespread functions than was initially thought.

same family (interleukin-8 and melanoma growth-stimulatory activity, for example) are known to be potent angiogenic factors¹¹. However, the physiological relevance of the angiogenic properties of these chemokines is still unclear, particularly as mice that lack the interleukin-8 receptor seem to have a normal vasculature¹².

The studies of Tachibana *et al.*³ and Zou *et al.*⁴ have revealed that CXCR4 is important in embryonic development, particularly in the formation of blood vessels in the gastrointestinal tract and in neuronal patterning within the CNS. The findings highlight new directions for chemokine research that biologists will need to consider, and they also make sense within the context of the biological properties of chemokines. As pointed out by Baggiolini², the chemoattractant properties of chemokines would be useful during morphogenesis to keep the cells that form tissues together. This is something that

chemokines do very well and, as we have seen with the CXCR4- and SDF-1-knockout mice, there are quite dramatic developmental consequences when these proteins are deleted.

Richard Horuk is in the Department of Immunology, Berlex Biosciences, Richmond, California 94804, USA.
e-mail: Richard_Horuk@berlex.com

1. Luster, A. D. *N. Engl. J. Med.* **338**, 436–445 (1998).
2. Baggiolini, M. *Nature* **392**, 565–568 (1998).
3. Tachibana, K. *et al. Nature* **393**, 591–594 (1998).
4. Zou, Y.-R., Kottmann, A. H., Kuroda, M., Taniuchi, I. & Littman, D. R. *Nature* **393**, 595–599 (1998).
5. Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. *Science* **272**, 872–877 (1996).
6. Bleul, C. C. *et al. Nature* **382**, 829–833 (1996).
7. Nagasawa, T. *et al. Nature* **382**, 635–638 (1996).
8. Richardson, R. M., Ali, H., Tomhave, E. D., Haribabu, B. & Snyderman, R. *J. Biol. Chem.* **270**, 27829–27833 (1995).
9. Hesselgesser, J. *et al. Curr. Biol.* **7**, 112–121 (1997).
10. Hesselgesser, J. *et al. Curr. Biol.* **8**, 595–598 (1998).
11. Moore, B. B., Arenberg, D. A. & Strieter, R. M. *Trends Cardiovasc. Med.* **8**, 51–58 (1998).
12. Cacalano, G. *et al. Science* **265**, 682–684 (1994).

Daedalus

Human quality control

Human reproduction is a highly uncertain business. Even trying hard, a woman may take many months to achieve pregnancy. Of her ova that manage to get fertilized, about 70% are likely to be miscarried — mostly so rapidly that she notices nothing but a missed or delayed period.

This, says Daedalus, shows evolution at work. Stable creatures like cats and cows can reproduce easily and unfailingly; but human beings are a recent and hasty evolutionary lash-up. Like a new generation of microprocessors, they push the state of the art so far that 100% yields cannot be expected. Flawed specimens are inevitable, and must be identified by rigorous quality-control tests. They are rejected rapidly, allowing their owner to try again.

So DREADCO biochemists are seeking to understand the 'dialogue' of tests between mother and fetus. It may be a strategic contest, like many other inter-generational genetic struggles; or it may be entirely honest. A set of genes trapped in a sub-standard fetus may prefer to be aborted, to give an identical set a better chance next time. One clue is that many drugs — among them alcohol, nicotine, caffeine and aminopterin — increase the spontaneous abortion rate. The conventional wisdom is that they may damage the fetus, so that it fails the mother's quality-control tests. On the other hand, they may sabotage the tests themselves, and trigger rejection by giving falsely low readings. Daedalus muses that they may even raise the tests' acceptance threshold. Mediocre fetuses, which might otherwise have just scraped an undistinguished pass, are then rejected; only the best are retained to term.

With luck, DREADCO's biochemical investigations and epidemiological studies will soon elucidate the testing process. Presumably the mother challenges the fetus with a chemical signal; the fetus releases, or fails to release, a competent hormonal response; and a chemical 'reject' signal triggers spontaneous abortion.

When identified, the 'reject' signal will be the ideal, natural, abortifacient. It will not damage the mother — evolution will have seen to that. It will be the perfect morning-after (or even month-after) contraceptive. And if it works by raising the acceptance level to some lofty peak of human perfection, then its occasional failure may be no bad thing. The mother will know that she is carrying an exceptional baby, well worth the inconvenience.

David Jones

Obituary

Derek H. R. Barton (1918–98)

Polymath of organic chemistry

With the death of Sir Derek Barton, on 16 March, we will miss not only a great chemical intellect but a fascinating and delightful human being. He was always full of plans and receptive to ideas, and he took pleasure in the company of his colleagues. Like his long-time colleague Geoffrey Wilkinson, he passed away suddenly and unexpectedly of a heart attack. In both cases this was the best way, as neither would have tolerated a state of forced inactivity.

Derek Barton was born in Gravesend, England, on 8 September 1918. His family were in what he referred to as “the wood business”. The young Barton followed in this for several years, but became acutely bored, and decided to seek a university degree, entering Imperial College, London, at 20. The late start was compensated by his securing a bachelor’s degree in only two years and then a PhD in only two more. He spent the next two years on war-related research and then another two in industry, where he did not “find the work very challenging in its intellectual content”. However, he was able to return to academia, first at Imperial College and then at Birkbeck, where he was appointed reader in 1950.

But the academic year before that was the crucial one of his career. He spent it at Harvard as a visiting lecturer, filling in for R. B. Woodward who was on sabbatical. Barton had already been thinking about the small barrier to rotation about the C–C bond in ethane, and about the work of Hassel on the shape (conformation, to use the technical term) of the cyclohexane ring. In a seminar given by the Harvard expert on steroid chemistry, L. F. Fieser, Barton proposed explanations for chemical facts that had not been explicable before, based on his developing ideas of ‘conformational analysis’. Fieser urged him to publish a short article expounding his explanation, and this article was the spark that lit the powder keg. Soon all organic chemists recognized that Barton’s conformational analysis held the key to understanding a multitude of problems (and in 1969 he was to receive the Nobel prize for this work, sharing it with Hassel).

His career moved rapidly thereafter. In 1957, after professorships at Birkbeck and Glasgow (where he received his FRS), he became professor of organic chemistry at Imperial College. There, Barton’s interests broadened into a great range of synthetic



problems, photochemistry (discovering the Barton reaction) and many other topics. He built up a stunning list of students and postdoctoral co-workers from all over the world who went on to be leaders in organic chemistry. He was also a tireless attendee of meetings and giver of lectures, and so became personally known and admired by virtually every important organic chemist in the world.

Along the way, his first marriage (from which there came his only child, a son) failed. At the same time his interest in and fondness for France grew, and led him to undertake earnest study of the language, which he did with the help of a charming French lady. Soon they were married, and he and Christiane made a very happy couple.

Nevertheless, a great black hole was inexorably approaching: retirement. The whole idea was abhorrent to him, and he averted it in a timely way. At the age of 59, in 1977, he became director of research at Gif-sur-Yvette, not far from Paris. With a French wife, a good position in France and an easy commute to Paris, Barton could indulge his Francophilia to the full. He also did more brilliant organic chemistry, and returned to what had always been a keen interest — radical chemistry — and to an activity he considered paramount — the invention of new reactions. The result of this was what he called ‘Gif chemistry’. Succinctly, this was a scheme to employ reagents based on iron to carry out selective oxidation

reactions on hydrocarbons.

Once again in 1985 the black hole of retirement loomed, and once again Barton was not prepared to retire. This brought him to Texas A&M. Why? Quite simply, because we offered him what he wanted: not a name chair, not a big office, not the role of being a brilliant but non-functional adornment. We offered a regular full professorship, a decent office and all the lab space he needed — which turned out to be a lot. Barton was thus able to continue his unflagging study of Gif chemistry and the invention of new reactions.

His time in College Station was both sad and happy. Not long after his arrival Christiane developed cancer, and after several very painful years for both, she died. Barton threw himself more than ever into his work, but he soon had the great good fortune to remarry (to Judith Cobb, a neighbour) and again enjoy a very well-rounded life.

One of Barton’s best friends, until his own untimely death in 1981, was R. B. Woodward. I was often struck by the similarities and differences between Barton and Woodward, probably the two most charismatic organic chemists of the postwar period. Each was both a quick study and a deep thinker. Each had an extraordinary memory, and each was a masterly lecturer, although in different styles. Each had the judgement to focus on big problems. Both enjoyed, until their respective ends, robust health whereby they could work extraordinarily hard. The main difference was that Barton led a well-managed life. In his love of fine wine, good food and, until his last few years, dry Martinis, Barton was second to none, but he was always in control. I offer this comparison because there was deep mutual respect between these two superstars, but also a spirit of friendly yet serious competition.

Barton’s ability to organize and concentrate his efforts was not the least of his talents. Although he was always generous with his time, he never wasted it. He was unfailingly collegial in entering into discussion of questions on which others sought advice, but it was never difficult to read the signs when he thought that enough had been said.

Barton’s death only a few months short of his eightieth birthday was a shock to all who knew and admired him. All that can be said is that it was a privilege to have had him with us for as long as we did.

F. Albert Cotton

*F. Albert Cotton is in the Department of Chemistry, Texas A&M University, College Station, Texas 77843-3255, USA.
e-mail: cotton@tamu.edu*

Mendeleev's matrix

Dmitri Mendeleev's periodic table permitted him to systematize crucial chemical data. But its real triumph was as an exercise in theoretical modelling, allowing the prediction of the discovery of previously unknown elements.

Martin Kemp

The 2D table is a wonderfully potent graphic device. Its very neatness bespeaks the achieving of precision, and it has proved its utility as an ordering and mnemonic system from the earliest documented eras of intellectual endeavour. Any table of 3×3 units or more embodies a surprising variety of potential orders — progressing in sequence from beginning to end, either in discontinuous rows or bucephadonically (by S-shaped motion), connected in individual columns, vertically, horizontally or diagonally, clustered with immediate neighbours, up to eight in number, and grouped in various blocks, lines or other configurations.

Some tables are matters of orderly convenience. Others graphically encode fundamental properties of natural systems. The king of all tables, ruling over the reconfiguring of the science of chemistry and governing much of its subsequent conduct, is Dmitri Mendeleev's periodic table of the elements of 1869.

As Mendeleev acknowledged, his table emerged from international endeavours to show that "the relations between the atomic weights... was governed by some general and simple laws". John Newlands's 1865 tabulation according to a musical "law of octaves" had been an inspiration much to Mendeleev's taste for all-embracing systems.

For his own part, Mendeleev seized upon the periodicity of elements arranged according to atomic weights and valencies, to provide the rules for what he called his game of "chemical solitaire". Each card in his deck was marked with the names or symbols of the elements, their weights and chemical

Tabelle I.

Typische Elemente			K = 39	Rb = 85	Cs = 133	—	—
			Ca = 40	Sr = 87	Ba = 137	—	—
H = 1	Li = 7	Na = 23	—	?Yt = 88?	?Di = 138?	Er = 178?	—
	Be = 9,4	Mg = 24	Ti = 48?	Zr = 90	Ce = 140?	?La = 180?	Tb = 231
B = 11	Al = 27,3	—	V = 51	Nb = 94	—	Ta = 182	—
C = 12	Si = 28	—	Cr = 52	Mo = 96	—	W = 184	U = 240
N = 14	P = 31	As = 75	Mn = 55	—	—	—	—
O = 16	S = 32	Se = 78	Fe = 56	Ru = 104	—	Os = 195?	—
F = 19	Cl = 35,5	Br = 80	Co = 59	Rh = 104	—	Ir = 197	—
			Ni = 59	Pd = 106	—	Pt = 198?	—
			Cu = 63	Ag = 108	—	Au = 199?	—
			Zn = 65	Cd = 112	—	Hg = 200	—
			—	In = 113	—	Tl = 204	—
			—	Sn = 118	—	Pb = 207	—
			—	Sb = 122	—	Bi = 208	—
			—	Te = 125?	—	—	—
			—	J = 127	—	—	—

der chemischen Elemente.

Mendeleev's periodic table, 1869.

properties — realizing the potential in Francis Bacon's "shuffle of things" in the most literal way.

The potency of Mendeleev's table was not just that it functioned as a tool for arranging properties but that the gaps in the sequences predicted "the discovery of yet unknown elements" (as first happened with gallium); that it became evident when an atomic weight may require emending; and that the properties of known and unknown elements could be predicted from their positions.

For Mendeleev, as a philosopher-chemist, who wrote on such things as the "unity of matter", the success of his table triumphantly proclaimed the value of what we

would call theoretical modelling in the face of narrow empiricism.

His Faraday Lecture in 1869 not only provided telling reviews of the rationale and development of his system but also delivered a powerful defence of conceptual structuring as a fundamental complement to experimental method. He constantly sought recourse to "agreement between theory and experiment; in other words, to demonstrated generalization and to the approved experiment".

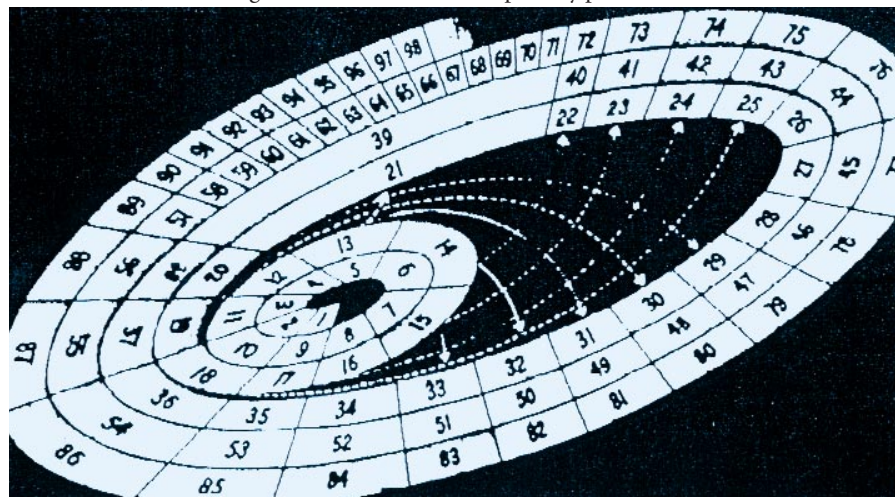
He proclaimed that "sound generalization — together with the relics of those which have proved to be untenable — promote scientific productivity, and ensure the luxurious growth of science under the influence of rays emanating from the centres of scientific energy". In his lecture, he told how "the inductive or experimental method of studying Nature gained a direct advantage from the old Pythagorean idea".

His tabular triumph stimulated a "luxurious growth" of alternative configurations, including wheels, flat spirals, helices, trees and intersecting vanes. When needed, as in the dynamic version devised for the Festival of Britain in 1951, the graphic rhetoric can be reworked in various styles, yet it is the rectilinear variety that has generally continued to reign supreme.

Justifiably, an icon depicting the periodic table was born aloft in Mendeleev's funeral procession in 1907 through the streets of St Petersburg.

Martin Kemp is in the Department of the History of Art, University of Oxford, 35 Beaumont Street, Oxford OX1 2PG, UK.

e-mail: martin.kemp@trinity.oxford.ac.uk



Festival of Britain periodic table, 1951.

Mauritian red nectar remains a mystery

Floral nectar is rich in chemicals and induces pollination¹. Although it may be tainted by algae or mould, it usually lacks colouring agents. However, a few plant species in Mauritius break this rule and produce red nectar. We attempted to find a function for this coloration, but its role remains unclear.

In 1976, a new species of plant in the family Campanulaceae, *Nesocodon mauritianus* (Fig. 1a), was found in Mauritius growing on inaccessible cliffs near the huge waterfall Cascade Cinq Cents Pied². This is the only known population and in 1997 we counted just 110–130 plants. So *N. mauritianus* is among the many endangered Mauritian species³. It was noted that the species' floral nectar was scarlet-red⁴ (Fig. 1a), an unusual characteristic that was subsequently overlooked.

We observed two bird species visiting flowers of *N. mauritianus*: the introduced red-whiskered bulbul (*Pycnonotus jocosus*) (Fig. 1c) and the native merle (*Hypsipetes olivaceus*). The bulbul pierced the base of the flower and robbed the nectar (80.8% of 52 flower visits, 40 hours of observation of 5 flowering plants), inserted its head into the flower to reach the nectar (9.6%) or tore the flower into pieces (9.6%). Only one merle was observed, and it picked an entire flower and flew away with it. No other animal was seen near the plants.

We mass-propagated *N. mauritianus* through seeds and analysed its nectar. We found that the nectar's pH was as high as 9.2 (the known pH range of all species is 3–10; ref. 1). When placed in acid, the red nectar turned yellow and showed antibacterial activity.

Using proton nuclear magnetic resonance spectroscopy, we identified the red pigment as a temperature-sensitive aurone (Fig. 1b). The yellow pigment was a light-sensitive 3-glucosylated flavone, or a flavonol with the same B-ring substitutions as the aurone. The corolla pigment was an anthocyanine with delphinidine as aglycon. All the pigments were products of flavonoid biosynthesis.

The nectar also contained 22–25% w/w hexose-dominated sugar. No specific floral scent was noticed, but hexanal, 2-hexenal, hexenol and hexanol were identified from leaves using gas-chromatography–mass spectrometry.

We examined other Mauritian flowers for red nectar and found it in *Trochetia blackburniana* (a species belonging to the family Sterculiaceae). In addition, *T. boutoniana* was reported locally⁵ to produce red nectar, and this was confirmed. We measured the sugar content of the latter as

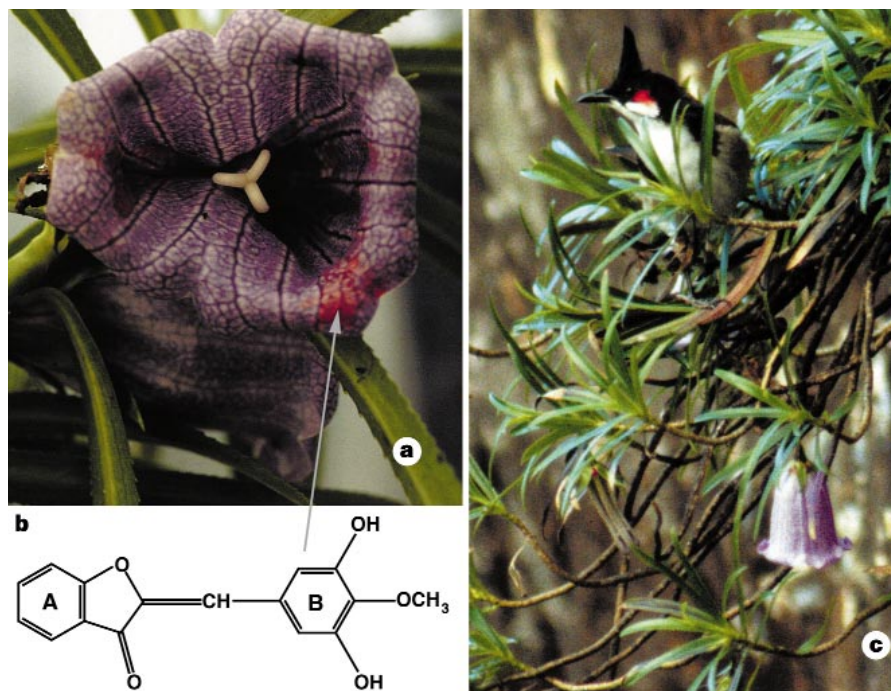


Figure 1 Red-coloured nectar in the Mauritian flower *Nesocodon mauritianus*. **a**, Red nectar flowing from a flower of *N. mauritianus*. **b**, Structure of the red-colouring aurone. **c**, *N. mauritianus* visited by a red-whiskered bulbul (*Pycnonotus jocosus*).

25–30% and found that its red pigment was an aurone too. *Trochetias* are pollinated and nectar-robbled by native white-eye birds (*Zosterops borbonicus* and *Z. chloronothos*)^{5–7}.

These three bird-pollinated Mauritian endemics are the only known plants in the world with coloured nectar. The red nectar in the blue flower of *N. mauritianus* may improve attractiveness to birds, but why do the red-flowered *Trochetias* produce red nectar?

We envisage at least three explanations for the evolution of this unique character: the original pollinator was an endemic but now extinct member of the Mauritian ‘afterlife’⁸ (recently extinct species); the red colour is an honest signal⁹ to pollinators improving their foraging efficiency and generating a fitness advantage over other bird-pollinated species; or the red pigment is associated with a deterrent against nectar robbers.

We regard the last two possibilities as unlikely. If red nectar is an honest signal, then why has it failed to evolve in other regions and taxa in which bird-pollination occurs? If it is a warning, then it is a poor ‘scarecrow’, as these taxa are heavily robbed of nectar.

We will search for extinct pollinators among recently extinct Mauritian species. However, ‘ghosts are not easy to pin down’⁸.

Jens M. Olesen

Department of Ecology & Genetics,
Aarhus University, Ny Munkegade Building 540,
DK-8000 Aarhus C, Denmark
e-mail: jens.olesen@biology.aau.dk

Nina Rønsted, Ulrik Tolderlund,

Claus Cornett, Per Mølgaard
Department of Pharmacognosy,
Royal Danish School of Pharmacy,
Universitetsparken 2,
DK-2100 Copenhagen Ø, Denmark

Jørn Madsen

The National Encyclopedia, Nytorv 17,
DK-1450 Copenhagen K, Denmark

Carl G. Jones

Jersey Wildlife Preservation Trust,
Black River, Republic of Mauritius

Carl E. Olsen

Chemistry Department,
Royal Danish Veterinary and Agricultural
University, Copenhagen, Thorvaldsensvej 40,
DK-1870 Frederiksberg C, Denmark

1. Baker, H. G. & Baker, I. in *Coevolution of Animals and Plants* (eds Gilbert, L. E. & Raven, P. H.) 100–140 (Univ. Texas Press, Austin, 1975).
2. Richardson, I. B. K. *Kew Bulletin* **33**, 547–550 (1979).
3. Wyse Jackson, P. S., Strahm, W., Cronk, Q. C. B. & Parnell, J. A. N. *Moorea* **7**, 35–45 (1988).
4. Wyse Jackson, P. S. *Kew Magazine* **7**, 113–117 (1990).
5. Staub, F. *Proc. R. Soc. Arts Sci. Mauritius* **5**, 7–78 (1987).
6. Safford, R. J. *Dodo. J. Jersey Wildlife Preserv. Trust* **27**, 113–138 (1987).
7. Staub, F. *Fauna of Mauritius and Associated Flora* (Indian Ocean, Rose-Hill, 1993).
8. Lawton, J. H. *Oikos* **73**, 145–147 (1995).
9. Zahavi, A. *J. Theor. Biol.* **53**, 205–214 (1975).

Moth uses fine tuning for odour resolution

Male moths, when responding to their species' blend of sex pheromones, cease their upwind flight when additional compounds are added to the mixture. Often these behavioural antagonists are the pheromone components of sympatric species that emit similar pheromone blends, and thus may function to prevent mating with females of the wrong species. Antagonists must be emitted from the same point source as the pheromone blend to be optimally effective¹⁻⁴, suggesting a fine discrimination between the occurrence of pheromone and antagonist.

Here we report that males of a North American noctuid moth species, *Helicoverpa zea*, can distinguish strands of pheromone from those of a behavioural antagonist separated by no more than 1 mm and, at most, 0.001 seconds.

The pheromone blend for *H. zea* is a 20:1 ratio of two components: (Z)-11-hexadecenal (Z11-16:Ald) and (Z)-9-hexadecenal (Z9-16:Ald). We found that male *H. zea* readily flew upwind and touched a Pasteur pipette source emitting this blend when it was pulsed at 5 s⁻¹ (with 0.02 s pulse duration; Fig. 1, treatments 1 and 3).

Significantly fewer males acted in this way when the antagonist (Z)-11-hexadecen-1-ol acetate (Z11-16:Ac) (ref. 5) was pulsed at the same frequency from the same pipette, the antagonist being added to a second filter paper at a loading of 50% of that of the pheromone (Fig. 1, treatment 7).

However, weaker antagonism resulted from placing this same amount of antagonist on a filter paper in a separate pipette whose tip was stationed only 1 mm either upwind or downwind of the tip of the pheromone-emitting pipette. Emission from this pipette was pulsed simultaneously with the pheromone (Fig. 1, treatment 5). Several other wind-tunnel experiments using similar protocols resulted in similar outcomes (results not shown).

Responses to other treatments suggested that the lack of optimal suppression of upwind flight to the non-coincident pulses was due to incomplete mixing of the strands of antagonist with those of the pheromone. The same two pipettes used in the pulsed treatment were made to emit continuously, allowing the plumes' two strands of antagonist and pheromone to mix more completely than during pulses as they travelled down the tunnel. Using this set-up, there was significantly greater antagonism of upwind flight (Fig. 1, treat-

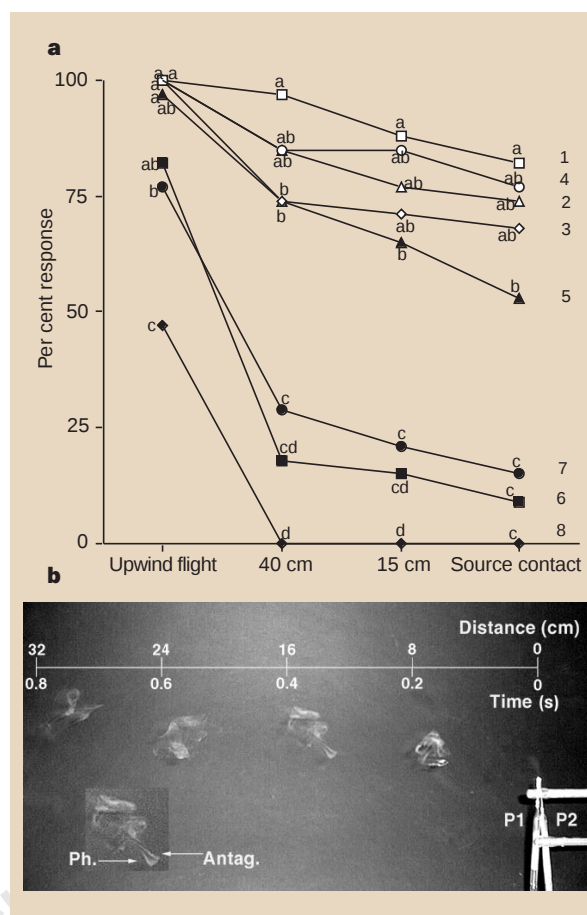


Figure 1 Responses of male moths to pheromone mixtures. **a**, Percentage of *H. zea* males flying upwind to pulsed or continuous plumes of pheromone alone (10 µg Z11-16:Ald + 0.5 µg Z9-16:Ald, always on one filter paper) or to plumes of pheromone also containing the antagonist (5 µg Z11-16:Ac, loaded onto a second filter paper and placed in the same or in a different pipette from the pheromone). Treatments are numbered 1-8 (right). Treatments 2 and 4 are pheromone-alone treatments using a continuous-emission (non-pulsed) regime. Percent responses in the same behavioural category having no letters in common are significantly different at $P < 0.05$; Chi-square 2 × 2 test of independence. **b**, Treatment 5, visualized by TiCl₄-generated smoke, consists of pulsed strands simultaneously generated at 5 s⁻¹ with 0.02 s duration and 5 ml s⁻¹ flow rate from two pipettes (P1 and P2) whose tips are separated by 1 mm. Ph., pheromone; antag., antagonist; arrows indicate incomplete mixing of the strands.

ment 6) than during pulsing (Fig. 1, treatment 5).

The intermittency of the signal was not in itself the cause of reduced antagonism. When the antagonist was loaded in the same pipette as the pheromone and pulsed (Fig. 1, treatment 7), there was as much antagonism as when the continuous regime was used. This was the case regardless of whether or not continuous emission occurred with the antagonist in the same pipette (Fig. 1, treatment 8) or in a separate pipette from the pheromone (Fig. 1, treatment 6).

Measurements of the ratios of pheromone to antagonist emitted from the pipette tips⁶ showed that the ratio during separate emissions was 198:1, and during co-emission from the same pipette was 225:1. It should be stressed that the incomplete mixing during pulsing would only separate the strands by 1 mm at most, or by 0.003 s if the male moth were stationary in the 40 cm s⁻¹ wind. Given the males' 90 cm s⁻¹ airspeed generated by the moth flying mainly, but not entirely, straight upwind, this temporal separation would be 0.001 s at most.

We propose that the olfactory processing system responsible for this remarkably fine olfactory resolution may arise from the organization of antennal neurons. In Lepidoptera, including *H. zea* (ref. 6), antennal neurons tuned to a known antagonist are

almost always co-compartmentalized within single receptor hairs with a neuron tuned to a pheromone component⁶⁻⁹. Only by sampling the air at the same point in space and time can a system integrating the inputs of two functionally different sensors, tuned to different compounds, determine whether there is complete spatial and temporal coincidence in the arrival of two odorants.

Such fine resolution could have evolved owing to its importance in mate finding. Males who could detect and respond to pure strands of pheromone, regardless of any imperfect mixing with antagonists, would have been at a selective mating advantage²⁻⁴.

T. C. Baker, H. Y. Fadamiro, A. A. Cosse

Department of Entomology, Iowa State University, Ames, Iowa 50011, USA

1. Rothschild, G. H. L. *Entomol. Exp. Appl.* **17**, 294-302 (1974).
2. Witzgall, P. & Priesner, E. *J. Chem. Ecol.* **17**, 1355-1362 (1991).
3. Liu, Y. B. & Haynes, K. F. *J. Chem. Ecol.* **18**, 299-307 (1992).
4. Liu, Y. B. & Haynes, K. F. *Physiol. Entomol.* **18**, 363-371 (1993).
5. Fadamiro, H. Y. & Baker, T. C. *Physiol. Entomol.* **22**, 316-324 (1997).
6. Cossé, A. A., Todd, J. L. & Baker, T. C. *J. Comp. Physiol. A* **182**, 585-594 (1988).
7. O'Connell, R. J. *et al. Science* **220**, 1408-1410 (1983).
8. Van der Pers, J. N. C. & Löfstedt, C. in *Mechanisms in Insect Olfaction* (eds Payne, T. L., Birch, M. C. & Kennedy, C. E. J.) 235-241 (Clarendon, Oxford, 1986).
9. Löfstedt, C., Hansson, B. S., Dijkerman, H. J., & Herrebut, W. *Physiol. Entomol.* **15**, 47-54 (1990).

Sound localization and neurons

Our ability to locate the direction of a sound's source depends largely on differences in the time that it takes for the sound to arrive at each ear. These interaural time delays (ITDs) can be extremely small (about 10 microseconds; ref. 1), so this ability is widely assumed to reflect the combined activity of many neurons^{2,3}. Fitzpatrick *et al.*⁴ estimated the pooling of 40 thalamic neurons to give a resolution of 16 microseconds. Our results, generated by a more conventional method, suggest that human thresholds might be determined by substantially fewer cells.

The ability of a neuron to discriminate between two stimuli is determined by the size of the difference in response relative to the response variability. The receiver operating characteristic⁵ analysis makes it possible to estimate this ability using the same metric as is used to quantify human psychophysical performance. This approach has previously been used to estimate the detection and discrimination of visual neurons^{6,7}.

We have applied receiver operating characteristic analysis to previously published results⁸ by assuming that response variability can be described by a Poisson distribution: $(\mu^x/x!)e^{-\mu}$, where μ denotes the mean. The percentage of correct responses increases with ITD (Fig. 1b,c).

In psychophysical experiments, it is conventional to define discrimination threshold as the stimulus magnitude at which the psychometric function crosses the 75% correct level. For the six neurons in the present study, the ITD at which this occurred ranged from 9.9 μ s up to 26.8 μ s (Fig. 1). The lowest of these values is comparable to the thresholds of human observers of about 10 microseconds (ref. 1).

Although these results are based on only six cells, it seems highly likely that there exist neurons with even lower thresholds. If one were to assume that an observer's discrimination threshold reflects the performance of the neuron with the lowest threshold, then one should have to compare the behavioural thresholds to that neuron. The psychophysical thresholds might therefore have been expected to be even lower than 10 microseconds.

These results indicate that single auditory neurons are able to distinguish ITDs as small as the human discrimination thresholds. But they are based on the assumption of a Poisson distribution. There are good reasons for expecting this kind of distribution for response variability^{9,10}, especially when only instantaneous variability is con-

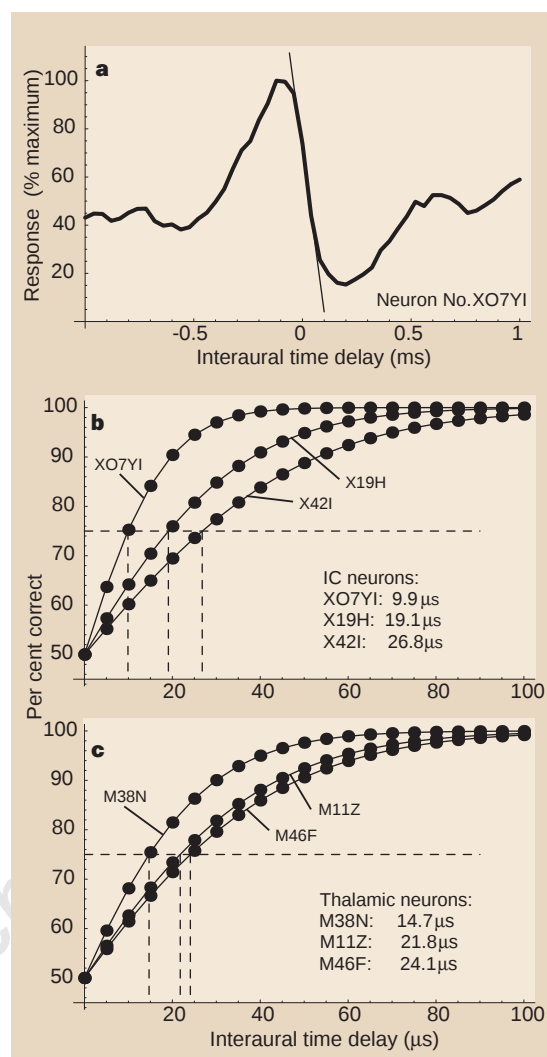


Figure 1 Analysis of interaural time delays (ITDs). **a**, Response against ITD plot of one inferior colliculus (IC) neuron (redrawn from Fig. 7 of ref. 8, neuron no. X07Y1). The very steep slope at zero ITD makes the neuron translate a small difference in ITD into a large difference in response. **b,c**, Percentage correct as a function of ITD for the three IC neurons (**b**) and three thalamic neurons (**c**). Horizontal broken lines, 75% correct level; vertical broken lines, ITD at which the functions cross the 75% correct level. The cell numbers of ref. 8 have been maintained to identify the neurons. The ITD corresponding to each neuron's 75% correct level is listed. One IC neuron (neuron no. X07Y1) has a discrimination threshold comparable to that of human observers.

sidered. It is known that response variability incorporates both an instantaneous and a long-term component^{6,7}. In sound localization, it is mainly the instantaneous variability that is relevant.

This might account for part of the large discrepancy between the present results and those of Fitzpatrick *et al.*⁴. That study combined neurons recorded at different times and from 31 different animals without differentiating long-term from instantaneous variability. This would result in an underestimate of the neurons' capabilities. Also, unlike the present analysis, they did not specifically exploit the steepest slopes of the response against ITD plots (Fig. 1a).

The present results should not be taken to mean that only one neuron is involved in sound localization. The responses of neurons are ambiguous. To remove this ambiguity, the combining of outputs from several neurons might be required. What my results do demonstrate is that it might not be necessary to pool the outputs from many neurons to account for the high accuracy with which human observers can localize sounds.

Bernt Christian Skottun

Skottun Research, 273 Mather Street,

Piedmont, California 94611-5154, USA

e-mail: bernt@best.com

1. Mills, A. W. *J. Acoust. Soc. Am.* **30**, 237–246 (1958).
2. Yin, T. C. T. & Chan, J. C. K. in *Auditory Function: Neurobiological Bases of Hearing* (eds Edelman, G. M., Gall, W. E. & Cowan, W. M.) 385–430 (Wiley, New York, 1988).
3. Hall, J. L. *J. Acoust. Soc. Am.* **37**, 814–823 (1965).
4. Fitzpatrick, D. C., Batra, R., Stanford, T. R. & Kudawa, S. *Nature* **388**, 871–874 (1997).
5. Green, D. M. & Swets, J. A. *Signal Detection Theory and Psychophysics* (Krieger, Huntington, NY, 1974).
6. Tolhurst, D. J., Movshon, J. A. & Dean, A. F. *Vision Res.* **23**, 775–785 (1983).
7. Bradley, A., Skottun, B. C., Ohzawa, I., Sclar, G. & Freeman, R. D. *J. Neurophysiol.* **57**, 755–772 (1987).
8. Stanford, T. R., Kuwada, S. & Batra, R. *J. Neurosci.* **12**, 3200–3216 (1992).
9. Rieke, F., Warland, D., de Ruyter van Steveninck, R. & Bialek, W. *Spikes* (MIT Press, 1997).
10. Tomko, G. J. & Crapper, D. R. *Brain Res.* **79**, 405–418 (1974).

scientific correspondence is intended to provide a forum for both brief, topical reports of general scientific interest and technical discussion of recently published material of particular interest to non-specialist readers. Priority will be given to contributions that have fewer than 500 words, 10 references and only one figure. Detailed guidelines are available on Nature's website or on request from nature@nature.com

Savage reversals

Rivers of Blood, Rivers of Gold: Europe's Conflict with Tribal Peoples

by Mark Cocker

Cape: 1998. Pp. 416. £20

Adam Kuper

The old school books used to tell the story of the European civilizing mission and the benighted savages who resisted it, embellished with the inspiring legends of Clive, Gordon or Livingstone. Mark Cocker's book tells very much the same story, "the confrontation between the civilised and savage", as he puts it. But he inverts the Victorian judgements and swaps around the heroes and villains, and instead of rousing accounts of Gordon being cut down by dervishes, he retells some of the goriest episodes of imperial brutality. The moral of the story is that the savages did not become civilized, but that the conquerors were turned into savages.

The bulk of the book is taken up by four stories, each dramatized as a confrontation between a vicious European leader and his tribal opponent. We have the conquest of Mexico in the early sixteenth century (Cortés versus Moctezuma), the destruction of the Tasmanians in the late nineteenth century (the British governor versus the Last Tasmanian), the more or less contemporary defeat of the Apache (Custer versus Geronimo), and the subjugation of Hendrik Witbooi's people by a German force in the colony of South West Africa in 1905.

Drawing on modern syntheses, without reference to detailed historical monographs, and relying entirely on sources in English, the case studies are highly sensationalized. But the main problem is Cocker's failure to discriminate and put them into context. He might have hinted at other, equally familiar but less bloody episodes, perhaps saying something about Hawaii or Buganda, or discussed Cook or Livingstone. He might even have made an effort to justify his conviction that European imperialists behaved very differently in the New World and the Old. (The Spanish conquest of the Americas is generally regarded, in important respects, as a continuation of the 'reconquest' of the Iberian Peninsula, which was associated with the repression of Moslems, Jews and unorthodox Catholics, and which provided a model for the repression of the Aztec and Inca civilizations.)

Cocker's Europeans are all chips off the same block, up to the same tricks over a period of 500 years. And all those who suffered from European colonialism during this long period, in four different continents, are also largely the same. He describes them unblushingly as savages (but in a loving, car-

ing way, as Dame Edna Everage would put it). And they are all, he says, tribal peoples. This is, of course, the stereotyped Victorian designation, evoking a picture of age-old ethnic groups, ruled by powerful chiefs, engaging in unspeakable rites under the direction of sinister magicians and priests, and constantly at war with their neighbours. If they have a fault, according to Cocker, it is this tendency to feuding, which makes them vulnerable to outside incursion.

The Europeans confronted very, very different peoples at different times and in different places. The Spanish stumbled on huge, well-organized, dazzlingly rich kingdoms, with literate élites and complex agricultural systems, at the hub of immense trading networks. The colonists in Australia were dealing with small, nomadic bands. The German soldiers in what is now Namibia came up against Dutch-speaking, gun-toting, often Christian 'coloured' or 'Baster' peoples, in many ways indistinguishable from the Boer frontiersmen who were causing the British such problems across the border, and who were regarded as enemies by many of the local Bantu-speaking peoples. The so-called

Apache were virtually bandit groups, established in the interstices between the British colonies and Mexico, and feared by their settled neighbours.

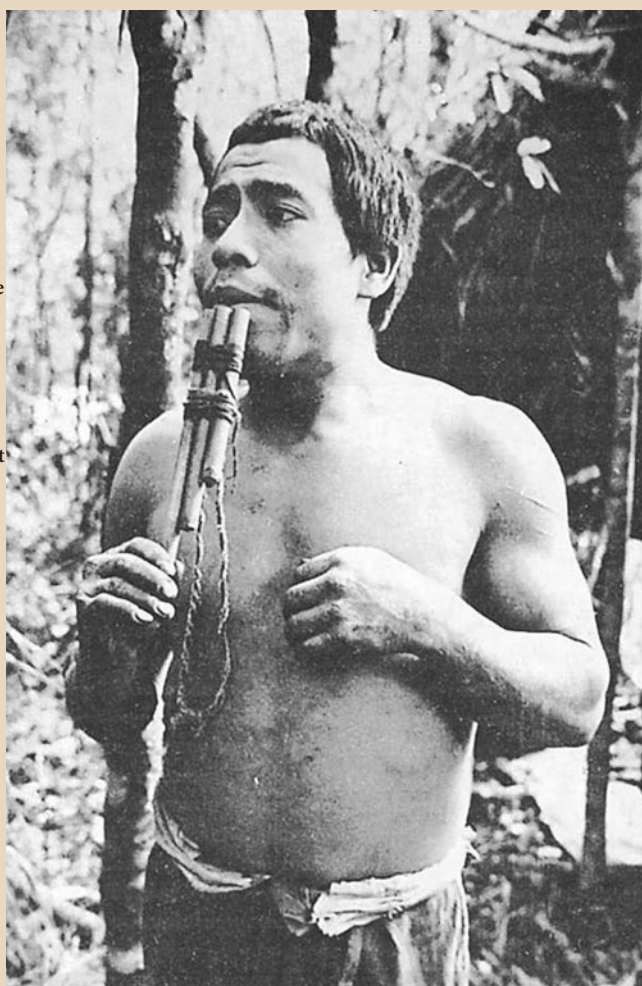
But for Cocker, the half a millennium of European imperialism is all the same, and its defining act is the bloody conquest of what he calls tribal or savage peoples. Moreover, as far as he is concerned, it is still going on. The French nuclear tests in the Mururoa Atoll represent just another episode in the history of the annihilation of native peoples.

Reading Cocker's book is a sobering experience. The writings of modern historians and anthropologists have obviously had little impact on popular mythologies. Even an enlightened journalist such as Cocker (he writes for the UK newspaper *The Guardian*) can write as though the world of Victorian scholarship was still intact. This book is just like the productions of Victorian imperialist propaganda, albeit with the signs reversed. It is the new morality, all right, but in the service of the old scholarship. □

Adam Kuper is in the Department of Social Anthropology, Brunel University, Uxbridge, Middlesex UB8 3PH, UK.

Lost cultures

In *Chronicle of the Guayaki Indians*, the French anthropologist Pierre Clastres gives an unsentimental yet sympathetic account of the year he spent with a little-known Paraguayan Indian tribe in 1963. First published in France in 1972 and here translated into English by the novelist Paul Auster, the book vividly describes the history, ritual, myths and culture of these now vanished people. Faber, £9.99 (pbk).



Eco-epidemiology

Health and Climate Change: Modelling the Impacts of Global Warming and Ozone Depletion

by Pim Martens

Earthscan: 1998. Pp. 176. £15 (pbk)

George A. Gellert

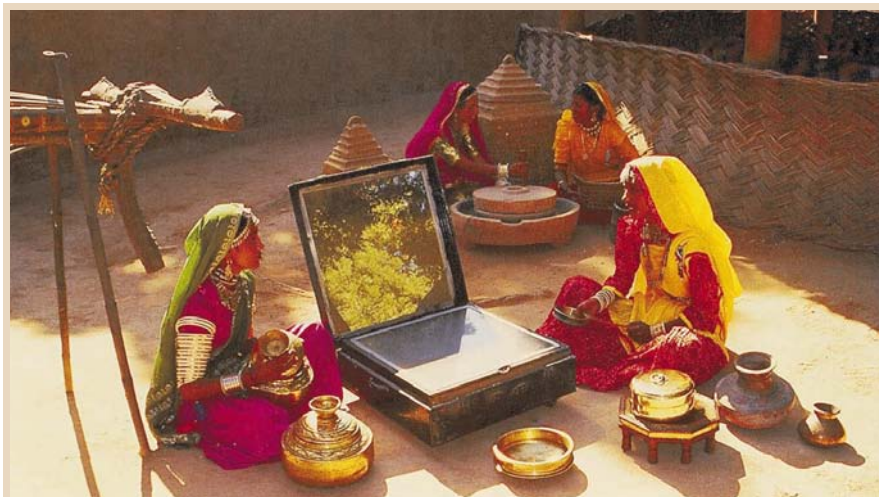
Modelling the effects of environmental health hazards presents scientists with unusual problems. With global climate change it is impossible to rely on empirical data as they are not readily available. In addition, even if non-alarmist projections are true, mankind may not have the time to risk standard prospective research on so potentially deleterious a set of health issues.

Pim Martens has provided a rigorous introduction to the purposes and practices of eco-epidemiology. The objective of this field is to increase competence in projecting the future effects of current trends in order to anticipate the negative effects of global atmospheric change. In developing such projections and estimates, new strategies and techniques are required to manage scientific uncertainty, and this is the focus of the book; the primary technique or tool it explores is simulation modelling. Martens sets out to look at the dynamics underlying the health effects of climate change and ozone depletion, along with their intrinsic uncertainties.

The rationale behind eco-epidemiological modelling is a need to provide a conceptual framework and analytical tools for evaluating situations and phenomena that extend beyond documented experience. Martens considers the development of multidisciplinary, integrated assessment models. His models incorporate projections of global environmental changes and extrapolations of the resulting effects on the health of individuals and populations derived from epidemiological data, biological knowledge, models of societal response, and historical experience. This brief and readable treatment is an introduction to the field and the modelling tools, and provides real-world illustrations of the use of the approach.

Martens begins by providing a succinct summary of the main differences between conventional epidemiological research methods and the systems-based assessment of global atmospheric changes. He then describes the main constructs of eco-epidemiological modelling, before moving quickly through several important areas of environmental threat.

Three potential consequences of global climate change are examined using eco-epidemiological modelling: shifts in the geography of vector-borne infectious diseases (malaria, dengue and schistosomiasis); alterations in exposure to thermal stress



Hard sell for the Sun

An Indian entrepreneur is trying to popularize a small solar cooker by hiring Rajasthani dancers to display it in the compound of a hotel. From *The Timeless Energy of the Sun* edited by

Madanjeet Singh (Thames and Hudson, £18.95). The book begins by describing how ancient peoples worshipped the Sun and goes on to describe the latest developments in solar power.

within urban populations and consequent thermal-related mortality (for example, cardiovascular and respiratory mortality); and increased incidence of skin cancer associated with increased ultraviolet levels resulting from ozone depletion. The progression of chapters and models builds added sensitivity and complexity into the method. For example, the second model presented on malaria refines the approach by including a simulation of adaptive processes.

Martens concludes with an attempt to translate the results of the methodology into useful information and a discussion of future lines of research.

In general, the book successfully achieves its objectives. One of its strengths is that it incorporates facets not usually covered in methodological reviews, such as a chapter providing estimates of the attributable population burdens of disease or mortality that may result from the global changes analysed. This information helps to make clear the critical link between the adverse effects modelled and their magnitude and specific health impacts.

But, if the book's breadth and effort to appeal to readers of diverse backgrounds and degrees of technical sophistication is a strength, it is also a weakness. The book seems to suffer an identity crisis, being neither introductory or interdisciplinary enough for the generalist, nor conceptually or methodologically deep enough for the expert. Is it a primer for the uninitiated or a case-studies reference for the practitioner? Nevertheless, this drawback is outweighed by a well organized and thoughtful presentation, along with an astute and effective use of graphics and tabular summaries. The book also makes extensive use of reference cita-

tions, but the topical index is so abbreviated as to have almost no value.

Martens recognizes the limitations of the models presented, including their immature stage of development, an inability to focus on local outcomes, and a lack of validation. He acknowledges the need for rapid evolution in the methodology, but points out that his case studies demonstrate the feasibility of such efforts to quantify systematically the health impacts of global climate change and ozone depletion. He has provided a solid contribution to an important, underexamined and scientifically complex discipline. □

George A. Gellert is at 1 Franklin Plaza, 200 North 16th Street, Philadelphia, PA 19101, USA.

Wherein blue genes?

Mood Genes: Hunting for Origins of Mania and Depression

by Samuel H. Barondes

W. H. Freeman: 1998. Pp. 237. \$24.95

George Fink

"Mania is a sickness for one's friends, depression for one's self. Both are chemical." This quotation from Robert Lowell sums up mood disorders, which are common and frequently lethal. According to the September 1997 report of the National Institute of Mental Health's Genetics Workgroup, chaired by Samuel H. Barondes, the lifetime risk of manic depression (bipolar disorder) is about 1% in the United States with no sex difference, whereas for a major depressive episode (unipolar disorder) the risk is 13% in men and 21% in women. Both illnesses contribute significantly to the high incidence of suicide.

Studies of twins show that the concordance rate for manic-depressive disorder is 80% in monozygotic twins compared with 8% in dizygotic twins. Together with adoption studies, these findings suggest that genetic factors are involved in the familial transmission of bipolar disorder. Although less compelling, the results of twin and adoption studies suggest that genetic factors are also involved in major depression. In *Mood Genes*, Barondes lucidly encompasses the development of modern psychiatry and genetics, and shows how the two are coupled in the quest for the genetic basis of mood disorders.

He starts with the story of Michael, a molecular biologist whose mother, Flora, is a writer with severe manic-depressive illness; she has bouts of profligate spending and sexual promiscuity followed by crashes into hopeless depression. When manic, she writes for at least 15 hours a day, producing 300 typed pages in just three weeks. Flora considers mania to be her "gemlike flame," reminiscent of the madness that drove many creative minds, as essayed by Kay Redfield Jamison in *Touched with Fire* (Simon and Schuster, 1993), which is a perfect complement to *Mood Genes*. Flora's brother, who is depressive and who has attempted suicide on several occasions, her father and Michael's son also exhibit signs consistent with manic-depressive illness. In his fifties, Michael feels down and takes the antidepressant Prozac, which throws him too into mania. It is on this family, affected by manic-depressive disorder in four successive generations, that Barondes builds the riveting story of psychiatry and genetics.

He explains how Emil Kraepelin tried to classify psychiatric disorders, as Rudolph Virchow had done for medical diseases. Kraepelin concluded from pedigree studies that manic-depressive insanity was a hereditary trait. To pursue this proposition he established in 1917 a Genealogical and Demographic Department in the Kaiser Wilhelm Institute for Psychiatry in Munich. But his work was undone by Ernst Rudin, appointed head of the department by Kraepelin, who was instrumental during the early Nazi era in drafting the infamous Law to Prevent Hereditarily Sick Offspring, which led to the compulsory sterilization of hundreds of thousands of people with real or constructed hereditary or congenital disease.

This tarnished Kraepelin's reputation and brought to an abrupt halt the development of psychiatric genetics, opening the way for Sigmund Freud's more compassionate, but arguably less rational, psychoanalytical approach to become the predominant force in psychiatry. The renaissance of interest in the classification of psychiatric disorders did not occur until the 1970s, forced in part by the development of powerful psychotropic drugs including lithium, the most effective drug for controlling manic symp-

toms. Preparation for the gene hunt is completed by the familiar account of the development of human genetics, from the peas of Gregor Mendel to the double helix and the genetic code.

The important concept of gene penetrance is introduced by reference to the genetically transmitted illness porphyria, which allegedly affected King George III and is expressed in only about one in ten carriers of the disease allele under conditions of 'biochemical overload'. Porphyria has heuristic value for understanding the genetic transmission of complex diseases, such as high blood pressure, diabetes and manic depression, in which factors extrinsic to the genome trigger the expression of the disease allele. This seems to have been the case for Michael, as he devel-

oped mania only after taking Prozac.

Barondes dismisses rather peremptorily the candidate gene approach (looking for genes that appear biologically relevant to the disorder) in the hunt for mood genes. Although it has found genes involved in other diseases, he says it has failed to find mood genes. But many studies are now in progress, and three recent reports link a polymorphism in the gene encoding the serotonin transporter, the target for Prozac, with mood disorders. Barondes instead focuses on the linkage approach, beginning his explanation with the studies of Thomas Morgan on the fruitfly, and underscoring the importance of restriction fragment length polymorphisms as gene markers.

James Gusella and Michael Conneally

Higher planes



In the words of internationally renowned photographer Emmet Gowin, the landscape "is always, in some sense, our home". For the past ten years, Gowin has used aerial photography to document man's mark on the Earth in images of industrial agriculture, mining explorations and nuclear test sites of the American West. In some sense, the aerial perspective provides the gift of seeing our own history, although known to us, for the first time. The exceptional beauty of the photographs is sometimes at odds with what they depict, and we are gently reminded that we are connected to, not separate from, these places. Gowin wishes simply to bear witness to

this reality, and his work is full of reverence for life. With the recent tragic halt to nuclear non-proliferation in Asia, let us hope that this philosophy finds global expression.

This photograph, "Subsidence Craters and the Yucca Fault, Northern End of Yucca Flat, Nevada Test Site", is part of an exhibition of Gowin's work taking place at the Pace Wildenstein MacGill Gallery in New York City until 13 June. A book of 250 of Gowin's photographs is planned.

Lisa Satterwhite is in the Department of Molecular Biology, Schultz Laboratory, Princeton University, Princeton, NJ 08544, USA.

discovered just such a polymorphism on chromosome 4, which is linked to Huntington's disease, after screening with only three probes. That this led to the identification of the genetic disease locus made linkage analysis seem relatively easy and highly effective. But the hunt for mood genes has so far proved to be a dismal failure. Claims made for linkages to the X chromosome and chromosomes 11 and 18, based on studies of Amish, Israeli and Costa Rican pedigrees respectively, have not been substantiated. Frustrated, Barondes turns to Alzheimer's disease, susceptibility to which is increased by the presence of the apolipoprotein A4 genotype, to reassure us that gene linkage can be useful in studying mental disorders.

Because the linkage approach has failed to identify mood genes, the peremptory dismissal of the candidate gene approach is puzzling. However, this and other inconsistencies do not detract from the overall importance of *Mood Genes* as an interim account of an exciting gene hunt, written in comfortable and in parts racy prose by an authority in the field. The evidence for the genetic transmission of mood disorders is incontrovertible, so the hunt must go on. □

George Fink is in the MRC Brain Metabolism Unit, Department of Pharmacology, University of Edinburgh, 1 George Square, Edinburgh EH8 9JZ, UK.

Animal anomalies

The Garden of Ediacara: Discovering the First Complex Life

by Mark A. S. McMenamin
Columbia University Press: 1998. Pp. 284.
\$29.95, £23.95

The Crucible of Creation: The Burgess Shale and the Rise of Animals

by Simon Conway Morris
Oxford University Press: 1998. Pp. 242.
£18.99, \$30

Douglas Palmer

Writing for a wide readership about obscure and extinct organisms without common names or modern analogues is peculiarly difficult. Nevertheless, the discovery of such fossils, especially when they are the remains of some of the earliest complex multicellular organisms, is so important that it warrants any serious attempt to tell their story. At least a book provides the scope to go beyond the constraints of the media soundbite.

The pitfalls of gross generalization and misrepresentation are only too well known. Recently, the main morning news programme on British radio raised the question of whether 'intelligent' life, as it was put, had evolved twice in the distant past. Furthermore, it was claimed that the first wave of 'innocent and peaceful, sentient beings' was

Past masters

Artist's impression of two resting *Styracosaurus albertensis*, taken from *The Horned Dinosaurs: A Natural History* by Peter Dodson, which is now out in paperback from Princeton University Press at \$19.95, £15.95. A "comprehensive" and "even-handed" work, wrote Angela Milner in a review of the original hardback edition (*Nature* 384, 426; 1996).



wiped out by the second wave, consisting of our ancestors, armed with tooth and buckler.

Behind the 'news' story is the publication of Mark McMenamin's *The Garden of Ediacara*, a popular account of his interpretation of the problematic soft-bodied organisms called the Ediacarans. He has based his ideas on those of the German palaeontologist Dolf Seilacher, who coined the term Vendobionta to describe what he sees as the curiously alien nature of these fossils as neither animals nor plants. Accordingly, the Ediacarans were a failed evolutionary experiment and were unceremoniously hustled off the stage of life by the advent of the more familiar Cambrian biota of arthropods and 'shelly' molluscs. Ediacarans have been found around the world, mainly in late Precambrian marine sediments dating from 600 million to 544 million years ago, but they are now known to have survived into Cambrian and possibly younger deposits.

It is a hard task to inform a non-specialist reader about the Ediacarans with their unusual mode of preservation in sandstones. The task is made even more difficult because there is now an intellectual free-for-all on the Ediacarans, unconstrained by much in the way of data. Ediacarans have variously been ascribed to lichens, protozoans, cnidarians, annelids and arthropods.

McMenamin is at his best when he has a good story to tell, such as his 1995 discovery of *Cyclomedusa* fossils from late Precambrian sediments 600 million years old in northern Mexico. A traditional tale, it has a sequence of trial and tribulation for the narrator 'hero' in his quest for the 'holy grail'. But too often his narrative approach as journalist, observer and confidant is not as satisfying as that adopted by Simon Conway Morris in *The Crucible of Creation*.

Conway Morris is concerned mainly with exploring the nature of the Burgess Shale organisms of middle Cambrian age, around 520 million years old. But his wide-ranging experience of other Cambrian deposits and

their biotas allows him to tell with authority a much broader story of their role in animal evolution. No doubt some academics' eyebrows will be raised by the intrusion of personal belief but, as much as anything else, this book is an account of an intellectual journey and not an academic paper. His authorial voice is a complex mix of omniscient narrator and guide/teacher, at times demanding or dismissive but always enormously informative and questioning.

Conway Morris has the advantage that the Burgess Shale fossils are now better defined than the Ediacarans, with plentiful, biologically based data and a good understanding of the taphonomy (*post mortem* processes of burial and fossilization) of the animals, both essential prerequisites for interpreting the fossil record. This has not always been so: as little as 20 years ago there were considerable misconceptions about some of the weird and wonderful creatures of the Burgess Shale, such as *Anomalocaris* and *Hallucigenia*.

When Stephen Jay Gould published his *Wonderful Life* in 1989, he claimed that so many of the forms of life present at that time were wiped out by extinctions that, if the tape of life were replayed, there would have been a very different set of survivors. Furthermore, chance would mitigate against humans being among them. Conway Morris argues against Gould's historical contingency, claiming that "it does not have such a meaningful effect on the totality of life" and "a different message can be read from the Burgess Shale".

Part of this argument extends back to the Ediacarans. Conway Morris makes a strong claim for at least one biological contact between the Ediacaran *Charniodiscus* and *Thaumaptilon*, a sea-pen from the Burgess. Perhaps the difficulties will be resolved only if other Ediacarans are found in a similar mode of preservation to the Burgess. □

Douglas Palmer is at 31 Mawson Road, Cambridge CB1 2DZ, UK.

Deciphering the biology of *Mycobacterium tuberculosis* from the complete genome sequence

S. T. Cole*, R. Brosch*, J. Parkhill, T. Garnier*, C. Churcher, D. Harris, S. V. Gordon*, K. Eiglmeier*, S. Gas*, C. E. Barry III†, F. Tekaia‡, K. Badcock, D. Basham, D. Brown, T. Chillingworth, R. Connor, R. Davies, K. Devlin, T. Feltwell, S. Gentles, N. Hamlin, S. Holroyd, T. Hornsby, K. Jagels, A. Krogh§, J. McLean, S. Moule, L. Murphy, K. Oliver, J. Osborne, M. A. Quail, M.-A. Rajandream, J. Rogers, S. Rutter, K. Seeger, J. Skelton, R. Squares, S. Squares, J. E. Sulston, K. Taylor, S. Whitehead & B. G. Barrell

Sanger Centre, Wellcome Trust Genome Campus, Hinxton CB10 1SA, UK

* Unité de Génétique Moléculaire Bactérienne, and ‡ Unité de Génétique Moléculaire des Levures, Institut Pasteur, 28 rue du Docteur Roux, 75724 Paris Cedex 15, France

† Tuberculosis Research Unit, Laboratory of Intracellular Parasites, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, Montana 59840, USA

§ Center for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark

Countless millions of people have died from tuberculosis, a chronic infectious disease caused by the tubercle bacillus. The complete genome sequence of the best-characterized strain of *Mycobacterium tuberculosis*, H37Rv, has been determined and analysed in order to improve our understanding of the biology of this slow-growing pathogen and to help the conception of new prophylactic and therapeutic interventions. The genome comprises 4,411,529 base pairs, contains around 4,000 genes, and has a very high guanine + cytosine content that is reflected in the biased amino-acid content of the proteins. *M. tuberculosis* differs radically from other bacteria in that a very large portion of its coding capacity is devoted to the production of enzymes involved in lipogenesis and lipolysis, and to two new families of glycine-rich proteins with a repetitive structure that may represent a source of antigenic variation.

Despite the availability of effective short-course chemotherapy (DOTS) and the Bacille Calmette-Guérin (BCG) vaccine, the tubercle bacillus continues to claim more lives than any other single infectious agent¹. Recent years have seen increased incidence of tuberculosis in both developing and industrialized countries, the widespread emergence of drug-resistant strains and a deadly synergy with the human immunodeficiency virus (HIV). In 1993, the gravity of the situation led the World Health Organisation (WHO) to declare tuberculosis a global emergency in an attempt to heighten public and political awareness. Radical measures are needed now to prevent the grim predictions of the WHO becoming reality. The combination of genomics and bioinformatics has the potential to generate the information and knowledge that will enable the conception and development of new therapies and interventions needed to treat this airborne disease and to elucidate the unusual biology of its aetiological agent, *Mycobacterium tuberculosis*.

The characteristic features of the tubercle bacillus include its slow growth, dormancy, complex cell envelope, intracellular pathogenesis and genetic homogeneity². The generation time of *M. tuberculosis*, in synthetic medium or infected animals, is typically ~24 hours. This contributes to the chronic nature of the disease, imposes lengthy treatment regimens and represents a formidable obstacle for researchers. The state of dormancy in which the bacillus remains quiescent within infected tissue may reflect metabolic shutdown resulting from the action of a cell-mediated immune response that can contain but not eradicate the infection. As immunity wanes, through ageing or immune suppression, the dormant bacteria reactivate, causing an outbreak of disease often many decades after the initial infection³. The molecular basis of dormancy and reactivation remains obscure but is expected to be genetically programmed and to involve intracellular signalling pathways.

The cell envelope of *M. tuberculosis*, a Gram-positive bacterium with a G + C-rich genome, contains an additional layer beyond the peptidoglycan that is exceptionally rich in unusual lipids, glycolipids and polysaccharides^{4,5}.

Novel biosynthetic pathways generate cell-wall components such as mycolic acids, mycosteric acid, phenolthiocerol, lipoarabinomannan and arabinogalactan, and several of these may contribute to mycobacterial longevity, trigger inflammatory host reactions and act in pathogenesis. Little is known about the mechanisms involved in life within the macrophage, or the extent and nature of the virulence factors produced by the bacillus and their contribution to disease.

It is thought that the progenitor of the *M. tuberculosis* complex, comprising *M. tuberculosis*, *M. bovis*, *M. bovis* BCG, *M. africanum* and *M. microti*, arose from a soil bacterium and that the human bacillus may have been derived from the bovine form following the domestication of cattle. The complex lacks interstrain genetic diversity, and nucleotide changes are very rare⁶. This is important in terms of immunity and vaccine development as most of the proteins will be identical in all strains and therefore antigenic drift will be restricted. On the basis of the systematic sequence analysis of 26 loci in a large number of independent isolates⁶, it was concluded that the genome of *M. tuberculosis* is either unusually inert or that the organism is relatively young in evolutionary terms.

Since its isolation in 1905, the H37Rv strain of *M. tuberculosis* has found extensive, worldwide application in biomedical research because it has retained full virulence in animal models of tuberculosis, unlike some clinical isolates; it is also susceptible to drugs and amenable to genetic manipulation. An integrated map of the 4.4 megabase (Mb) circular chromosome of this slow-growing pathogen had been established previously and ordered libraries of cosmids and bacterial artificial chromosomes (BACs) were available^{7,8}.

Organization and sequence of the genome

Sequence analysis. To obtain the contiguous genome sequence, a combined approach was used that involved the systematic sequence analysis of selected large-insert clones (cosmids and BACs) as well as

random small-insert clones from a whole-genome shotgun library. This culminated in a composite sequence of 4,411,529 base pairs (bp) (Figs 1, 2), with a G + C content of 65.6%. This represents the second-largest bacterial genome sequence currently available (after that of *Escherichia coli*)⁹. The initiation codon for the *dnaA* gene, a hallmark for the origin of replication, *oriC*, was chosen as the start point for numbering. The genome is rich in repetitive DNA, particularly insertion sequences, and in new multigene families and duplicated housekeeping genes. The G + C content is relatively constant throughout the genome (Fig. 1) indicating that horizontally transferred pathogenicity islands of atypical base composition are probably absent. Several regions showing higher than average G + C content (Fig. 1) were detected; these correspond to sequences belonging to a large gene family that includes the polymorphic G + C-rich sequences (PGRSs).

Genes for stable RNA. Fifty genes coding for functional RNA molecules were found. These molecules were the three species produced by the unique ribosomal RNA operon, the 16S rRNA involved in degradation of proteins encoded by abnormal messenger RNA, the RNA component of RNase P, and 45 transfer RNAs. No 4.5S RNA could be detected. The *rrn* operon is situated unusually as it occurs about 1,500 kilobases (kb) from the putative *oriC*; most eubacteria have one or more *rrn* operons near to *oriC* to exploit the gene-dosage effect obtained during replication¹⁰. This arrangement may be related to the slow growth of *M. tuberculosis*. The genes encoding tRNAs that recognize 43 of the 61 possible sense codons were distributed throughout the genome and, with one

exception, none of these uses A in the first position of the anticodon, indicating that extensive wobble occurs during translation. This is consistent with the high G + C content of the genome and the consequent bias in codon usage. Three genes encoding tRNAs for methionine were found; one of these genes (*metV*) is situated in a region that may correspond to the terminus of replication (Figs 1, 2). As *metV* is linked to defective genes for integrase and excisionase, perhaps it was once part of a phage or similar mobile genetic element.

Insertion sequences and prophages. Sixteen copies of the promiscuous insertion sequence IS6110 and six copies of the more stable element IS1081 reside within the genome of H37Rv⁸. One copy of IS1081 is truncated. Scrutiny of the genomic sequence led to the identification of a further 32 different insertion sequence elements, most of which have not been described previously, and of the 13E12 family of repetitive sequences which exhibit some of the characteristics of mobile genetic elements (Fig. 1). The newly discovered insertion sequences belong mainly to the IS3 and IS256 families, although six of them define a new group. There is extensive similarity between IS1561 and IS1552 with insertion sequence elements found in *Nocardia* and *Rhodococcus* spp., suggesting that they may be widely disseminated among the actinomycetes.

Most of the insertion sequences in *M. tuberculosis* H37Rv appear to have inserted in intergenic or non-coding regions, often near tRNA genes (Fig. 1). Many are clustered, suggesting the existence of insertional hot-spots that prevent genes from being inactivated, as has been described for *Rhizobium*¹¹. The chromosomal distribution of the insertion sequences is informative as there appears to have been a selection against insertions in the quadrant encompassing *oriC* and an overrepresentation in the direct repeat region that contains the prototype IS6110. This bias was also observed experimentally in a transposon mutagenesis study¹².

At least two prophages have been detected in the genome sequence and their presence may explain why *M. tuberculosis* shows persistent low-level lysis in culture. Prophages phiRv1 and phiRv2 are both ~10 kb in length and are similarly organized, and some of their gene products show marked similarity to those encoded by certain bacteriophages from *Streptomyces* and saprophytic mycobacteria. The site of insertion of phiRv1 is intriguing as it corresponds to part of a repetitive sequence of the 13E12 family that itself appears to have integrated into the biotin operon. Some strains of *M. tuberculosis* have been described as requiring biotin as a growth supplement, indicating either that phiRv1 has a polar effect on expression of the distal *bio* genes or that aberrant excision, leading to mutation, may occur. During the serial attenuation of *M. bovis* that led to the vaccine strain *M. bovis* BCG, the phiRv1 prophage was lost¹³. In a systematic study of the genomic diversity of prophages and insertion sequences (S.V.G. *et al.*, manuscript in preparation), only IS1532 exhibited significant variability, indicating that most of the prophages and insertion sequences are currently stable. However, from these combined observations, one can conclude that horizontal transfer of genetic material into the free-living ancestor of the *M. tuberculosis* complex probably occurred in nature before the tubercle bacillus adopted its specialized intracellular niche.

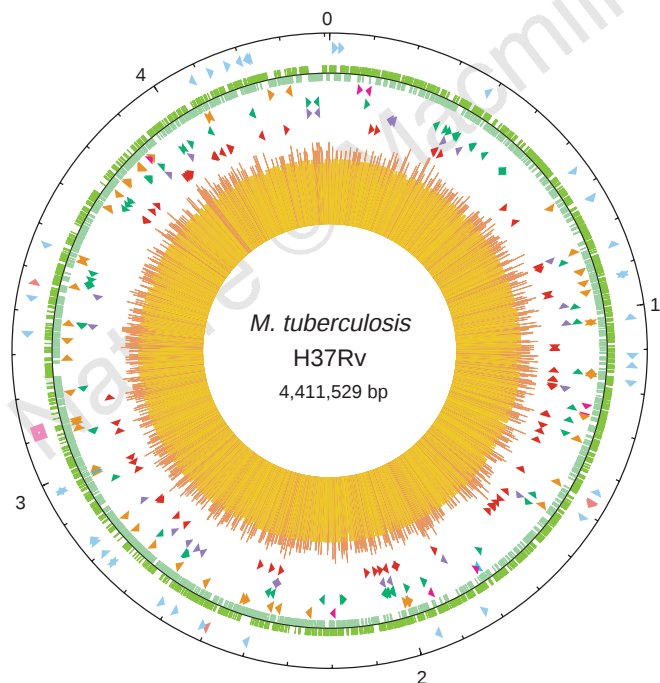


Figure 1 Circular map of the chromosome of *M. tuberculosis* H37Rv. The outer circle shows the scale in Mb, with 0 representing the origin of replication. The first ring from the exterior denotes the positions of stable RNA genes (tRNAs are blue, others are pink) and the direct repeat region (pink cube); the second ring inwards shows the coding sequence by strand (clockwise, dark green; anticlockwise, light green); the third ring depicts repetitive DNA (insertion sequences, orange; 13E12 REP family, dark pink; prophage, blue); the fourth ring shows the positions of the PPE family members (green); the fifth ring shows the PE family members (purple, excluding PGRS); and the sixth ring shows the positions of the PGRS sequences (dark red). The histogram (centre) represents G + C content, with <65% G + C in yellow, and >65% G + C in red. The figure was generated with software from DNASTAR.

Figure 2 Linear map of the chromosome of *M. tuberculosis* H37Rv showing the position and orientation of known genes and coding sequences (CDS). We used the following functional categories (adapted from ref. 20): lipid metabolism (black); intermediary metabolism and respiration (yellow); information pathways (pink); regulatory proteins (sky blue); conserved hypothetical proteins (orange); proteins of unknown function (light green); insertion sequences and phage-related functions (blue); stable RNAs (purple); cell wall and cell processes (dark green); PE and PPE protein families (magenta); virulence, detoxification and adaptation (white). For additional information about gene functions, refer to <http://www.sanger.ac.uk>.

Genes encoding proteins. 3,924 open reading frames were identified in the genome (see Methods), accounting for ~91% of the potential coding capacity (Figs 1, 2). A few of these genes appear to have in-frame stop codons or frameshift mutations (irrespective of the source of the DNA sequenced) and may either use frameshifting during translation or correspond to pseudogenes. Consistent with the high G + C content of the genome, GTG initiation codons (35%) are used more frequently than in *Bacillus subtilis* (9%) and *E. coli* (14%), although ATG (61%) is the most common translational start. There are a few examples of atypical initiation codons, the most notable being the ATC used by *infC*, which begins with ATT in both *B. subtilis* and *E. coli*^{9,14}. There is a slight bias in the orientation of the genes (Fig. 1) with respect to the direction of replication as ~59% are transcribed with the same polarity as replication, compared with 75% in *B. subtilis*. In other bacteria, genes transcribed in the same direction as the replication forks are believed to be expressed more efficiently^{9,14}. Again, the more even distribution in gene polarity seen in *M. tuberculosis* may reflect the slow growth and infrequent replication cycles. Three genes (*dnaB*, *recA* and *Rv1461*) have been invaded by sequences encoding inteins (protein introns) and in all three cases their counterparts in *M. leprae* also contain inteins, but at different sites¹⁵ (S.T.C. *et al.*, unpublished observations).

Protein function, composition and duplication. By using various database comparisons, we attributed precise functions to ~40% of the predicted proteins and found some information or similarity for another 44%. The remaining 16% resembled no known proteins and may account for specific mycobacterial functions. Examination of the amino-acid composition of the *M. tuberculosis* proteome by correspondence analysis¹⁶, and comparison with that of other microorganisms whose genome sequences are available, revealed a statistically significant preference for the amino acids Ala, Gly, Pro, Arg and Trp, which are all encoded by G + C-rich codons, and a comparative reduction in the use of amino acids encoded by A + T-rich codons such as Asn, Ile, Lys, Phe and Tyr (Fig. 3). This approach also identified two groups of proteins rich in Asn or Gly that belong to new families, PE and PPE (see below). The fraction of the proteome that has arisen through gene duplication is similar to that seen in *E. coli* or *B. subtilis* (~51%; refs 9, 14), except that the level of sequence conservation is considerably higher, indicating that there may be extensive redundancy or differential production of the corresponding polypeptides. The apparent lack of divergence following gene duplication is consistent with the hypothesis that *M. tuberculosis* is of recent descent⁶.

General metabolism, regulation and drug resistance

Metabolic pathways. From the genome sequence, it is clear that the tubercle bacillus has the potential to synthesize all the essential amino acids, vitamins and enzyme co-factors, although some of the pathways involved may differ from those found in other bacteria. *M. tuberculosis* can metabolize a variety of carbohydrates, hydrocarbons, alcohols, ketones and carboxylic acids^{2,17}. It is apparent from genome inspection that, in addition to many functions involved in lipid metabolism, the enzymes necessary for glycolysis, the pentose phosphate pathway, and the tricarboxylic acid and glyoxylate cycles are all present. A large number (~200) of oxidoreductases, oxygenases and dehydrogenases is predicted, as well as many oxygenases containing cytochrome P450, that are similar to fungal proteins involved in sterol degradation. Under aerobic growth conditions, ATP will be generated by oxidative phosphorylation from electron transport chains involving a ubiquinone cytochrome *b* reductase complex and cytochrome *c* oxidase. Components of several anaerobic phosphorylative electron transport chains are also present, including genes for nitrate reductase (*narGHJI*), fumarate reductase (*frdABCD*) and possibly nitrite reductase (*nirBD*), as well as a new reductase (*narX*) that results from a rearrangement of a homologue of the *narGHJI* operon. Two genes encoding haemoglobin-like

proteins, which may protect against oxidative stress or be involved in oxygen capture, were found. The ability of the bacillus to adapt its metabolism to environmental change is significant as it not only has to compete with the lung for oxygen but must also adapt to the microaerophilic/anaerobic environment at the heart of the burgeoning granuloma.

Regulation and signal transduction. Given the complexity of the environmental and metabolic choices facing *M. tuberculosis*, an extensive regulatory repertoire was expected. Thirteen putative sigma factors govern gene expression at the level of transcription initiation, and more than 100 regulatory proteins are predicted (Table 1). Unlike *B. subtilis* and *E. coli*, in which there are >30 copies of different two-component regulatory systems¹⁴, *M. tuberculosis* has only 11 complete pairs of sensor histidine kinases and response regulators, and a few isolated kinase and regulatory genes. This relative paucity in environmental signal transduction pathways is probably offset by the presence of a family of eukaryotic-like serine/threonine protein kinases (STPKs), which function as part of a phosphorelay system¹⁸. The STPKs probably have two domains: the well-conserved kinase domain at the amino terminus is predicted to be connected by a transmembrane segment to the carboxy-terminal region that may respond to specific stimuli. Several of the predicted envelope lipoproteins, such as that encoded by *lppR* (Rv2403), show extensive similarity to this putative receptor domain of STPKs, suggesting possible interplay. The STPKs probably function in signal transduction pathways and may govern important cellular decisions such as dormancy and cell division, and although their partners are unknown, candidate genes for phosphoprotein phosphatases have been identified.

Drug resistance. *M. tuberculosis* is naturally resistant to many antibiotics, making treatment difficult¹⁹. This resistance is due mainly to the highly hydrophobic cell envelope acting as a permeability barrier⁴, but many potential resistance determinants are also encoded in the genome. These include hydrolytic or drug-modifying enzymes such as β -lactamases and aminoglycoside acetyl transferases, and many potential drug-efflux systems, such as 14 members of the major facilitator family and numerous ABC transporters. Knowledge of these putative resistance mechanisms will promote better use of existing drugs and facilitate the conception of new therapies.

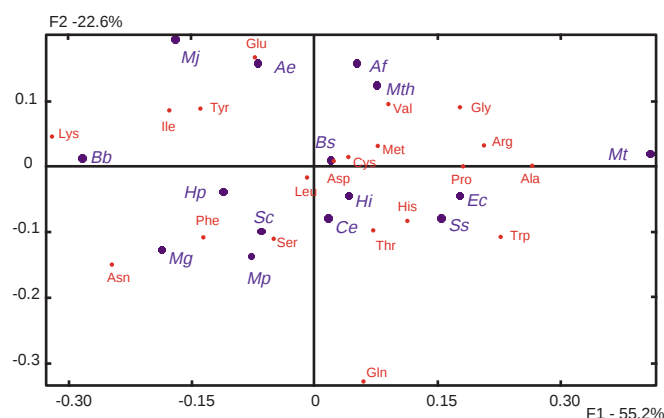


Figure 3 Correspondence analysis of the proteomes from extensively sequenced organisms as a function of amino-acid composition. Note the extreme position of *M. tuberculosis* and the shift in amino-acid preference reflecting increasing G + C content from left to right. Abbreviations used: Ae, *Aquifex aeolicus*; Af, *Archaeoglobus fulgidis*; Bb, *Borrelia burgdorferi*; Bs, *B. subtilis*; Ce, *Caenorhabditis elegans*; Ec, *E. coli*; Hi, *Haemophilus influenzae*; Hp, *Helicobacter pylori*; Mg, *Mycoplasmma genitalium*; Mj, *Methanococcus jannaschii*; Mp, *Mycoplasmma pneumoniae*; Mt, *M. tuberculosis*; Mth, *Methanobacterium thermoautotrophicum*; Sc, *Saccharomyces cerevisiae*; Ss, *Synechocystis* sp. strain PCC6803. F1 and F2, first and second factorial axes¹⁶.

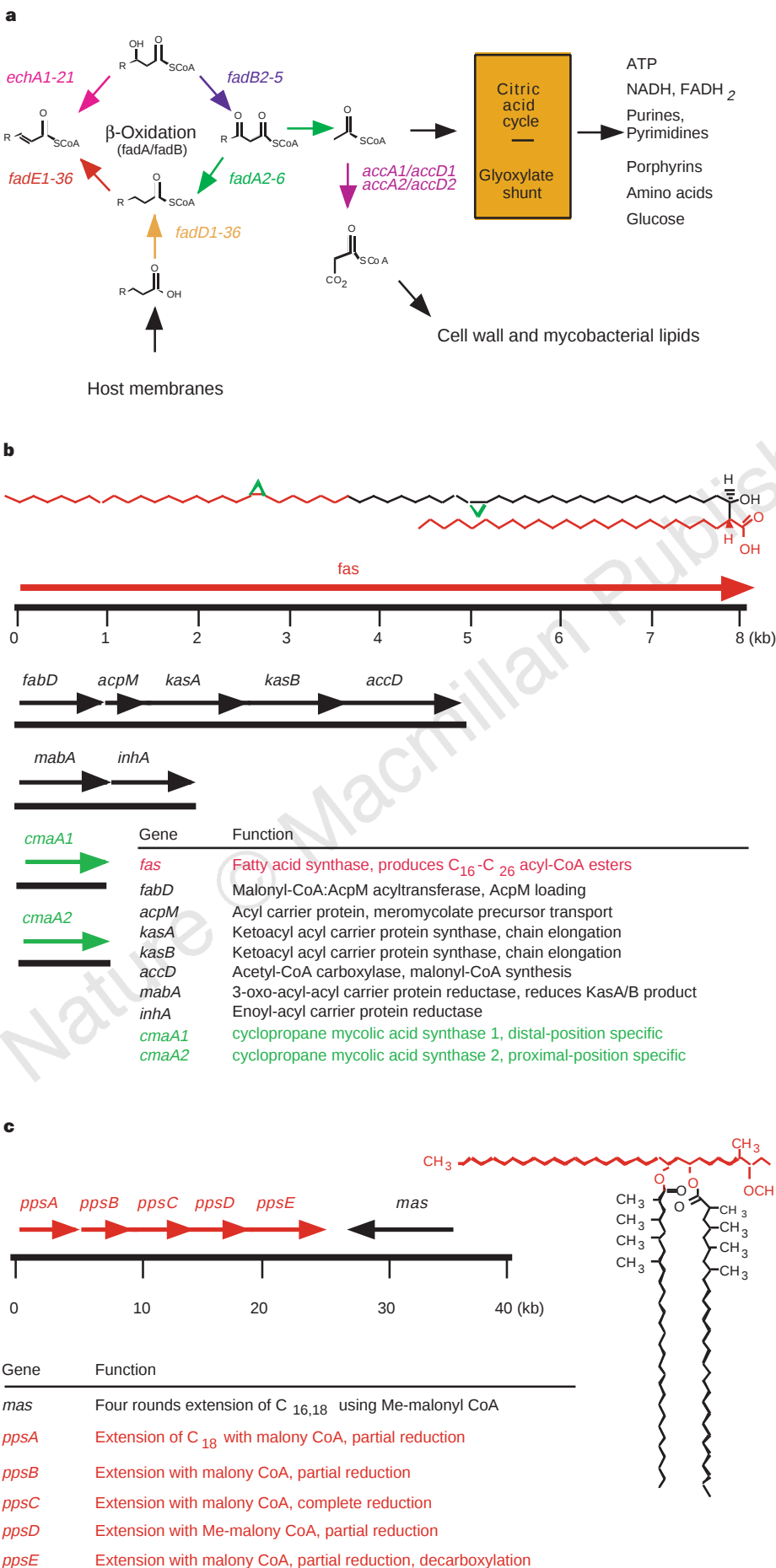


Figure 4 Lipid metabolism. **a**, Degradation of host-cell lipids is vital in the intracellular life of *M. tuberculosis*. Host-cell membranes provide precursors for many metabolic processes, as well as potential precursors of mycobacterial cell-wall constituents, through the actions of a broad family of β -oxidative enzymes encoded by multiple copies in the genome. These enzymes produce acetyl CoA, which can be converted into many different metabolites and fuel for the bacteria through the actions of the enzymes of the citric acid cycle and the glyoxylate shunt of this cycle. **b**, The genes that synthesize mycolic acids, the dominant lipid component of the mycobacterial cell wall, include the type I fatty acid synthase (*fas*) and a unique type II system which relies on extension of a precursor bound to an acyl carrier protein to form full-length (~80-carbon) mycolic acids. The *cma* genes are responsible for cyclopropanation. **c**, The genes that produce phthiocerol dimycocerosate form a large operon and represent type I (*mas*) and type II (the *pps* operon) polyketide synthase systems. Functions are colour coordinated.

Very few organisms produce such a diverse array of lipophilic molecules as *M. tuberculosis*. These molecules range from simple fatty acids such as palmitate and tuberculostearate, through isoprenoids, to very-long-chain, highly complex molecules such as mycolic acids and the phenolphthiocerol alcohols that esterify with mycoserolic acid to form the scaffold for attachment of the mycosides. Mycobacteria contain examples of every known lipid and polyketide biosynthetic system, including enzymes usually found in mammals and plants as well as the common bacterial systems. The biosynthetic capacity is overshadowed by the even more remarkable radiation of degradative, fatty acid oxidation systems and, in total, there are ~250 distinct enzymes involved in fatty acid metabolism in *M. tuberculosis* compared with only 50 in *E. coli*²⁰.

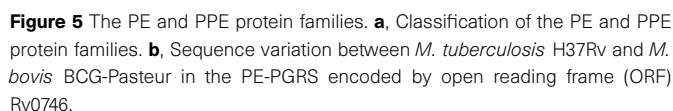
because they are difficult to distinguish from other members of the short-chain alcohol dehydrogenase family on the basis of primary sequence. The five enzymes that complete the cycle by thiolysis of the β -ketoester, the acetyl-CoA C-acetyltransferases, do indeed appear to be a more limited family. In addition to this extensive set of dissociated degradative enzymes, the genome also encodes the canonical FadA/FadB β -oxidation complex (Rv0859 and Rv0860). Accessory activities are present for the metabolism of odd-chain and multiply unsaturated fatty acids.

Fatty acid biosynthesis. At least two discrete types of enzyme system, fatty acid synthase (FAS) I and FAS II, are involved in fatty acid biosynthesis in mycobacteria (Fig. 4b). FAS I (Rv2524, *fas*) is a single polypeptide with multiple catalytic activities that generates several shorter CoA esters from acetyl-CoA primers⁵ and probably creates precursors for elongation by all of the other fatty acid and polyketide systems. FAS II consists of dissociable enzyme components which act on a substrate bound to an acyl-carrier protein (ACP). FAS II is incapable of *de novo* fatty acid synthesis but instead elongates palmitoyl-ACP to fatty acids ranging from 24 to 56 carbons in length^{17,21}. Several different components of FAS II may be targets for the important tuberculosis drug isoniazid, including the enoyl-ACP reductase *InhA*²², the ketoacyl-ACP synthase *KasA* and the ACP *AcpM*²¹. Analysis of the genome shows that there are only three potential ketoacyl synthases: *KasA* and *KasB* are highly related, and their genes cluster with *acpM*, whereas *KasC* is a more distant homologue of a ketoacyl synthase III system. The number of ketoacyl synthase and ACP genes indicates that there is a single FAS II system. Its genetic organization, with two clustered ketoacyl synthases, resembles that of type II aromatic polyketide biosynthetic gene clusters, such as those for actinorhodin, tetracycline and tetracenomycin in *Streptomyces* species²³. *InhA* seems to be the sole enoyl-ACP reductase and its gene is co-transcribed with a *fabG* homologue, which encodes 3-oxoacyl-CoA reductase. Both of these proteins are probably important in the biosynthesis of mycolic acids.

Fatty acids are synthesized from malonyl-CoA and precursors are generated by the enzymatic carboxylation of acetyl (or propionyl)-CoA by a biotin-dependent carboxylase (Fig. 4b). From study of the genome we predict that there are three complete carboxylase systems, each consisting of an α - and a β -subunit, as well as three β -subunits without an α -counterpart. As a group, all of the carboxylases seem to be more related to the mammalian homologues than to the corresponding bacterial enzymes. Two of these carboxylase systems (*accA1*, *accD1* and *accA2*, *accD2*) are probably involved in degradation of odd-numbered fatty acids, as they are adjacent to genes for other known degradative enzymes. They may convert propionyl-CoA to succinyl-CoA, which can then be incorporated into the tricarboxylic acid cycle. The synthetic carboxylases (*accA3*, *accD3*, *accD4*, *accD5* and *accD6*) are more difficult to understand. The three extra β -subunits might direct carboxylation to the appropriate precursor or may simply increase the total amount of carboxylated precursor available if this step were rate-limiting.

Synthesis of the paraffinic backbone of fatty and mycolic acids in the cell is followed by extensive postsynthetic modifications and unsaturations, particularly in the case of the mycolic acids^{24,25}. Unsaturation is catalysed either by a FabA-like β -hydroxyacyl-ACP dehydrase, acting with a specific ketoacyl synthase, or by an aerobic terminal mixed function desaturase that uses both molecular oxygen and NADPH. Inspection of the genome revealed no obvious candidates for the FabA-like activity. However, three potential aerobic desaturases (encoded by *desA1*, *desA2* and *desA3*) were evident that show little similarity to related vertebrate or yeast enzymes (which act on CoA esters) but instead resemble plant desaturases (which use ACP esters). Consequently, the genomic data indicate that unsaturation of the meromycolate chain may occur while the acyl group is bound to AcpM.

Much of the subsequent structural diversity in mycolic acids is



generated by a family of *S*-adenosyl-L-methionine-dependent enzymes, which use the unsaturated meromycolic acid as a substrate to generate *cis* and *trans* cyclopropanes and other mycolates. Six members of this family have been identified and characterized²⁵ and two clustered, convergently transcribed new genes are evident in the genome (*umaA1* and *umaA2*). From the functions of the known family members and the structures of mycolic acids in *M. tuberculosis*, it is tempting to speculate that these new enzymes may introduce the *trans* cyclopropanes into the meromycolate precursor. In addition to these two methyltransferases, there are two other unrelated lipid methyltransferases (*Ufa1* and *Ufa2*) that share homology with cyclopropane fatty acid synthase of *E. coli*²⁵. Although cyclopropanation seems to be a relatively common modification of mycolic acids, cyclopropanation of plasma-membrane constituents has not been described in mycobacteria. Tuberculoheptanoic acid is produced by methylation of oleic acid, and may be synthesized by one of these two enzymes.

Condensation of the fully functionalized and preformed meromycolate chain with a 26-carbon α -branch generates full-length mycolic acids that must be transported to their final location for attachment to the cell-wall arabinogalactan. The transfer and subsequent transesterification is mediated by three well-known immunogenic proteins of the antigen 85 complex²⁶. The genome encodes a fourth member of this complex, antigen 85C' (*fbpC2*, *Rv0129*), which is highly related to antigen 85C. Further studies are needed to show whether the protein possesses mycolyltransferase activity and to clarify the reason behind the apparent redundancy.

Polyketide synthesis. Mycobacteria synthesize polyketides by several different mechanisms. A modular type I system, similar to that involved in erythromycin biosynthesis²³, is encoded by a very large operon, *ppsABCDE*, and functions in the production of phenolphthiocerol⁵. The absence of a second type I polyketide synthase suggests that the related lipids phthiocerol A and B, phthiodiolone A and phthiotriol may all be synthesized by the same system, either from alternative primers or by differential postsynthetic modification. It is physiologically significant that the *pps* gene cluster occurs immediately upstream of *mas*, which encodes the multifunctional enzyme mycocerosic acid synthase (MAS), as their products phthiocerol and mycocerosic acid esterify to form the very abundant cell-wall-associated molecule phthiocerol dimycocerosate (Fig. 4c).

Members of another large group of polyketide synthase enzymes are similar to MAS, which also generates the multiply methyl-branched fatty acid components of mycosides and phthiocerol dimycocerosate, abundant cell-wall-associated molecules⁵. Although some of these polyketide synthases may extend type I FAS CoA primers to produce other long-chain methyl-branched fatty acids such as mycolipenic, mycolipodienic and mycolipanic acids or the phthioceranic and hydroxyphthioceranic acids, or may even show functional overlap⁵, there are many more of these enzymes than there are known metabolites. Thus there may be new lipid and polyketide metabolites that are expressed only under certain conditions, such as during infection and disease.

A fourth class of polyketide synthases is related to the plant enzyme superfamily that includes chalcone and stilbene synthase²³. These polyketide synthases are phylogenetically divergent from all other polyketide and fatty acid synthases and generate unreduced polyketides that are typically associated with anthocyanin pigments and flavonoids. The function of these systems, which are often linked to apparent type I modules, is unknown. An example is the gene cluster spanning *pk10*, *pk7*, *pk8* and *pk9*, which includes two of the chalcone-synthase-like enzymes and two modules of an apparent type I system. The unknown metabolites produced by these enzymes are interesting because of the potent biological activities of some polyketides such as the immunosuppressor rapamycin.

Siderophores. Peptides that are not ribosomally synthesized are

made by a process that is mechanistically analogous to polyketide synthesis^{23,27}. These peptides include the structurally related iron-scavenging siderophores, the mycobactins and the exochelins^{2,28}, which are derived from salicylate by the addition of serine (or threonine), two lysines and various fatty acids and possible polyketide segments. The *mbt* operon, encoding one apparent salicylate-activating protein, three amino-acid ligases, and a single module of a type I polyketide synthase, may be responsible for the biosynthesis of the mycobacterial siderophores. The presence of only one non-ribosomal peptide-synthesis system indicates that this pathway may generate both siderophores and that subsequent modification of a single ϵ -amino group of one lysine residue may account for the different physical properties and function of the siderophores²⁸.

Immunological aspects and pathogenicity

Given the scale of the global tuberculosis burden, vaccination is not only a priority but remains the only realistic public health intervention that is likely to affect both the incidence and the prevalence of the disease²⁹. Several areas of vaccine development are promising, including DNA vaccination, use of secreted or surface-exposed proteins as immunogens, recombinant forms of BCG and rational attenuation of *M. tuberculosis*²⁹. All of these avenues of research will benefit from the genome sequence as its availability will stimulate more focused approaches. Genes encoding ~90 lipoproteins were identified, some of which are enzymes or components of transport systems, and a similar number of genes encoding preproteins (with type I signal peptides) that are probably exported by the Sec-dependent pathway. *M. tuberculosis* seems to have two copies of *secA*. The potent T-cell antigen Esat-6 (ref. 30), which is probably secreted in a Sec-independent manner, is encoded by a member of a multigene family. Examination of the genetic context reveals several similarly organized operons that include genes encoding large ATP-hydrolysing membrane proteins that might act as transporters. One of the surprises of the genome project was the discovery of two extensive families of novel glycine-rich proteins, which may be of immunological significance as they are predicted to be abundant and potentially polymorphic antigens.

The PE and PPE multigene families. About 10% of the coding capacity of the genome is devoted to two large unrelated families of acidic, glycine-rich proteins, the PE and PPE families, whose genes are clustered (Figs 1, 2) and are often based on multiple copies of the polymorphic repetitive sequences referred to as PGRSs, and major polymorphic tandem repeats (MPTRs), respectively^{31,32}. The names PE and PPE derive from the motifs Pro-Glu (PE) and Pro-Pro-Glu (PPE) found near the N terminus in most cases³³. The 99 members of the PE protein family all have a highly conserved N-terminal domain of ~110 amino-acid residues that is predicted to have a globular structure, followed by a C-terminal segment that varies in size, sequence and repeat copy number (Fig. 5). Phylogenetic analysis separated the PE family into several subfamilies. The largest of these is the highly repetitive PGRS class, which contains 61 members; members of the other subfamilies, share very limited sequence similarity in their C-terminal domains (Fig. 5). The predicted molecular weights of the PE proteins vary considerably as a few members contain only the N-terminal domain, whereas most have C-terminal extensions ranging in size from 100 to 1,400 residues. The PGRS proteins have a high glycine content (up to 50%), which is the result of multiple tandem repetitions of Gly-Gly-Ala or Gly-Gly-Asn motifs, or variations thereof.

The 68 members of the PPE protein family (Fig. 5) also have a conserved N-terminal domain that comprises ~180 amino-acid residues, followed by C-terminal segments that vary markedly in sequence and length. These proteins fall into at least three groups, one of which constitutes the MPTR class characterized by the presence of multiple, tandem copies of the motif Asn-X-Gly-X-Gly-Asn-X-Gly. The second subgroup contains a characteristic, well-conserved motif around position 350, whereas the third contains

proteins that are unrelated except for the presence of the common 180-residue PPE domain.

The subcellular location of the PE and PPE proteins is unknown and in only one case, that of a lipase (Rv3097), has a function been demonstrated. On examination of the protein database from the extensively sequenced *M. leprae*¹⁵, no PGRS- or MPTR-related polypeptides were detected but a few proteins belonging to the non-MPTR subgroup of the PPE family were found. These proteins include one of the major antigens recognized by leprosy patients, the serine-rich antigen³⁴. Although it is too early to attribute biological functions to the PE and PPE families, it is tempting to speculate that they could be of immunological importance. Two interesting possibilities spring to mind. First, they could represent the principal source of antigenic variation in what is otherwise a genetically and antigenically homogeneous bacterium. Second, these glycine-rich proteins might interfere with immune responses by inhibiting antigen processing.

Several observations and results support the possibility of antigenic variation associated with both the PE and the PPE family proteins. The PGRS member Rv1759 is a fibronectin-binding protein of relative molecular mass 55,000 (ref. 35) that elicits a variable antibody response, indicating either that individuals mount different immune responses or that this PGRS protein may vary between strains of *M. tuberculosis*. The latter possibility is supported by restriction fragment length polymorphisms for various PGRS and MPTR sequences in clinical isolates³³. Direct support for genetic variation within both the PE and the PPE families was obtained by comparative DNA sequence analysis (Fig. 5). The gene for the PE-PGRS protein Rv0746 of BCG differs from that in H37Rv by the deletion of 29 codons and the insertion of 46 codons. Similar variation was seen in the gene for the PPE protein Rv0442 (data not shown). As these differences were all associated with repetitive sequences they could have resulted from intergenic or intragenic recombinational events or, more probably, from strand slippage during replication³². These mechanisms are known to generate antigenic variability in other bacterial pathogens³⁶.

There are several parallels between the PGRS proteins and the Epstein-Barr virus nuclear antigens (EBNAs). Members of both polypeptide families are glycine-rich, contain extensive Gly-Ala repeats, and exhibit variation in the length of the repeat region between different isolates. The Gly-Ala repeat region of EBNA1 functions as a *cis*-acting inhibitor of the ubiquitin/proteasome antigen-processing pathway that generates peptides presented in the context of major histocompatibility complex (MHC) class I molecules^{37,38}. MHC class I knockout mice are very susceptible to *M. tuberculosis*, underlining the importance of a cytotoxic T-cell response in protection against disease^{3,39}. Given the many potential effects of the PPE and PE proteins, it is important that further studies are performed to understand their activity. If extensive antigenic variability or reduced antigen presentation were indeed found, this would be significant for vaccine design and for understanding protective immunity in tuberculosis, and might even explain the varied responses seen in different BCG vaccination programmes⁴⁰.

Pathogenicity. Despite intensive research efforts, there is little information about the molecular basis of mycobacterial virulence⁴¹. However, this situation should now change as the genome sequence will accelerate the study of pathogenesis as never before, because other bacterial factors that may contribute to virulence are becoming apparent. Before the completion of the genome sequence, only three virulence factors had been described⁴¹: catalase-peroxidase, which protects against reactive oxygen species produced by the phagocyte; *mce*, which encodes macrophage-colonizing factor⁴²; and a sigma factor gene, *sigA* (aka *rpoV*), mutations in which can lead to attenuation⁴¹. In addition to these single-gene virulence factors, the mycobacterial cell wall⁴ is also important in pathology,

but the complex nature of its biosynthesis makes it difficult to identify critical genes whose inactivation would lead to attenuation.

On inspection of the genome sequence, it was apparent that four copies of *mce* were present and that these were all situated in operons, comprising eight genes, organized in exactly the same manner. In each case, the genes preceding *mce* code for integral membrane proteins, whereas *mce* and the following five genes are all predicted to encode proteins with signal sequences or hydrophobic stretches at the N terminus. These sets of proteins, about which little is known, may well be secreted or surface-exposed; this is consistent with the proposed role of Mce in invasion of host cells⁴². Furthermore, a homologue of *smpB*, which has been implicated in intracellular survival of *Salmonella typhimurium*, has also been identified⁴³. Among the other secreted proteins identified from the genome sequence that could act as virulence factors are a series of phospholipases C, lipases and esterases, which might attack cellular or vacuolar membranes, as well as several proteases. One of these phospholipases acts as a contact-dependent haemolysin (N. Stoker, personal communication). The presence of storage proteins in the bacillus, such as the haemoglobin-like oxygen captors described above, points to its ability to stockpile essential growth factors, allowing it to persist in the nutrient-limited environment of the phagosome. In this regard, the ferritin-like proteins, encoded by *bfrA* and *bfrB*, may be important in intracellular survival as the capacity to acquire enough iron in the vacuole is very limited. □

Methods

Sequence analysis. Initially, ~3.2 Mb of sequence was generated from cosmids⁸ and the remainder was obtained from selected BAC clones⁷ and 45,000 whole-genome shotgun clones. Sheared fragments (1.4–2.0 kb) from cosmids and BACs were cloned into M13 vectors, whereas genomic DNA was cloned in pUC18 to obtain both forward and reverse reads. The PGRS genes were grossly underrepresented in pUC18 but better covered in the BAC and cosmid M13 libraries. We used small-insert libraries⁴⁴ to sequence regions prone to compression or deletion and, in some cases, obtained sequences from products of the polymerase chain reaction or directly from BACs⁷. All shotgun sequencing was performed with standard dye terminators to minimize compression problems, whereas finishing reactions used dRhodamine or BigDye terminators (<http://www.sanger.ac.uk>). Problem areas were verified by using dye primers. Thirty differences were found between the genomic shotgun sequences and the cosmids; twenty of which were due to sequencing errors and ten to mutations in cosmids (1 error per 320 kb). Less than 0.1% of the sequence was from areas of single-clone coverage, and <0.2% was from one strand with only one sequencing chemistry.

Informatics. Sequence assembly involved PHRAP, GAP4 (ref. 45) and a customized perl script that merges sequences from different libraries and generates segments that can be processed by several finishers simultaneously. Sequence analysis and annotation was managed by DIANA (B.G.B. *et al.*, unpublished). Genes encoding proteins were identified by TB-parse⁴⁶ using a hidden Markov model trained on known *M. tuberculosis* coding and non-coding regions and translation-initiation signals, with corroboration by positional base preference. Interrogation of the EMBL, TrEMBL, SwissProt, PROSITE⁴⁷ and in-house databases involved BLASTN, BLASTX⁴⁸, DOTTER (<http://www.sanger.ac.uk>) and FASTA⁴⁹. tRNA genes were located and identified using tRNAscan and tRNAscan-SE⁵⁰. The complete sequence, a list of annotated cosmids and linking regions can be found on our website (<http://www.sanger.ac.uk>) and in MycDB (<http://www.pasteur.fr/mycdb/>).

Received 15 April; accepted 8 May 1998.

1. Snider, D. E. Jr, Ravighione, M. & Kochi, A. in *Tuberculosis: Pathogenesis, Protection, and Control* (ed. Bloom, B. R.) 2–11 (Am. Soc. Microbiol., Washington DC, 1994).
2. Wheeler, P. R. & Ratledge, C. in *Tuberculosis: Pathogenesis, Protection, and Control* (ed. Bloom, B. R.) 353–385 (Am. Soc. Microbiol., Washington DC, 1994).
3. Chan, J. & Kaufmann, S. H. E. in *Tuberculosis: Pathogenesis, Protection, and Control* (ed. Bloom, B. R.) 271–284 (Am. Soc. Microbiol., Washington DC, 1994).
4. Brennan, P. J. & Draper, P. in *Tuberculosis: Pathogenesis, Protection, and Control* (ed. Bloom, B. R.) 271–284 (Am. Soc. Microbiol., Washington DC, 1994).
5. Kolattukudy, P. E., Fernandes, N. D., Azad, A. K., Fitzmaurice, A. M. & Sirakova, T. D. Biochemistry

- and molecular genetics of cell-wall lipid biosynthesis in mycobacteria. *Mol. Microbiol.* **24**, 263–270 (1997).
6. Sreevatsan, S. *et al.* Restricted structural gene polymorphism in the *Mycobacterium tuberculosis* complex indicates evolutionarily recent global dissemination. *Proc. Natl Acad. Sci. USA* **94**, 9869–9874 (1997).
 7. Brosch, R. *et al.* Use of a *Mycobacterium tuberculosis* H37Rv bacterial artificial chromosome library for genome mapping, sequencing and comparative genomics. *Infect. Immun.* **66**, 2221–2229 (1998).
 8. Philipp, W. J. *et al.* An integrated map of the genome of the tubercle bacillus, *Mycobacterium tuberculosis* H37Rv, and comparison with *Mycobacterium leprae*. *Proc. Natl Acad. Sci. USA* **93**, 3132–3137 (1996).
 9. Blattner, F. R. *et al.* The complete genome sequence of *Escherichia coli* K-12. *Science* **277**, 1453–1462 (1997).
 10. Cole, S. T. & Saint-Girons, I. Bacterial genomics. *FEMS Microbiol. Rev.* **14**, 139–160 (1994).
 11. Freiberg, C. *et al.* Molecular basis of symbiosis between *Rhizobium* and legumes. *Nature* **387**, 394–401 (1997).
 12. Bardarov, S. *et al.* Conditionally replicating mycobacteriophages: a system for transposon delivery to *Mycobacterium tuberculosis*. *Proc. Natl Acad. Sci. USA* **94**, 10961–10966 (1997).
 13. Mahairas, G. G., Sabo, P. J., Hickey, M. J., Singh, D. C. & Stover, C. K. Molecular analysis of genetic differences between *Mycobacterium bovis* BCG and virulent *M. bovis*. *J. Bacteriol.* **178**, 1274–1282 (1996).
 14. Kunst, F. *et al.* The complete genome sequence of the gram-positive bacterium *Bacillus subtilis*. *Nature* **390**, 249–256 (1997).
 15. Smith, D. R. *et al.* Multiplex sequencing of 1.5 Mb of the *Mycobacterium leprae* genome. *Genome Res.* **7**, 802–819 (1997).
 16. Greenacre, M. *Theory and Application of Correspondence Analysis* (Academic, London, 1984).
 17. Ratledge, C. R. in *The Biology of the Mycobacteria* (eds Ratledge, C. & Stanford, J.) 53–94 (Academic, San Diego, 1982).
 18. Av-Gay, Y. & Davies, J. Components of eukaryotic-like protein signaling pathways in *Mycobacterium tuberculosis*. *Microb. Comp. Genomics* **2**, 63–73 (1997).
 19. Cole, S. T. & Telenti, A. Drug resistance in *Mycobacterium tuberculosis*. *Eur. Resp. Rev.* **8**, 701S–713S (1995).
 20. Riley, M. & Labeledan, B. in *Escherichia coli* and *Salmonella* (ed. Neidhardt, F. C.) 2118–2202 (ASM, Washington, 1996).
 21. Mdluli, K. *et al.* Inhibition of a *Mycobacterium tuberculosis* β -ketoacyl ACP synthase by isoniazid. *Science* **280**, 1607–1610 (1998).
 22. Banerjee, A. *et al.* *inhA*, a gene encoding a target for isoniazid and ethionamide in *Mycobacterium tuberculosis*. *Science* **263**, 227–230 (1994).
 23. Hopwood, D. A. Genetic contributions to understanding polyketide synthases. *Chem. Rev.* **97**, 2465–2497 (1997).
 24. Minnikin, D. E. in *The Biology of the Mycobacteria* (eds Ratledge, C. & Stanford, J.) 95–184 (Academic, London, 1982).
 25. Barry, C. E. III *et al.* Mycolic acids: structure, biosynthesis, and physiological functions. *Prog. Lipid Res.* (in the press).
 26. Belisle, J. T. *et al.* Role of the major antigen of *Mycobacterium tuberculosis* in cell wall biogenesis. *Science* **276**, 1420–1422 (1997).
 27. Marahiel, M. A., Stachelhaus, T. & Mootz, H. D. Modular peptide synthetases involved in nonribosomal peptide synthesis. *Chem. Rev.* **97**, 2651–2673 (1997).
 28. Gobin, J. *et al.* Iron acquisition by *Mycobacterium tuberculosis*: isolation and characterization of a family of iron-binding exochelins. *Proc. Natl Acad. Sci. USA* **92**, 5189–5193 (1995).
 29. Young, D. B. & Fruth, U. in *New Generation Vaccines* (eds Levine, M., Woodrow, G., Kaper, J. & Cobon, G. S.) 631–645 (Marcel Dekker, New York, 1997).
 30. Sorensen, A. L., Nagai, S., Houen, G., Andersen, P. & Anderson, A. B. Purification and characterization of a low-molecular-mass T-cell antigen secreted by *Mycobacterium tuberculosis*. *Infect. Immun.* **63**, 1710–1717 (1995).
 31. Hermans, P. W. M., van Soolingen, D. & van Embden, J. D. A. Characterization of a major polymorphic tandem repeat in *Mycobacterium tuberculosis* and its potential use in the epidemiology of *Mycobacterium kansasii* and *Mycobacterium goodii*. *J. Bacteriol.* **174**, 4157–4165 (1992).
 32. Poulet, S. & Cole, S. T. Characterisation of the polymorphic GC-rich repetitive sequence (PGRS) present in *Mycobacterium tuberculosis*. *Arch. Microbiol.* **163**, 87–95 (1995).
 33. Cole, S. T. & Barrell, B. G. in *Genetics and Tuberculosis* (eds Chadwick, D. J. & Cardew, G., Novartis Foundation Symp. 217) 160–172 (Wiley, Chichester, 1998).
 34. Vega-Lopez, F. *et al.* Sequence and immunological characterization of a serine-rich antigen from *Mycobacterium leprae*. *Infect. Immun.* **61**, 2145–2153 (1993).
 35. Abou-Zeid, C. *et al.* Genetic and immunological analysis of *Mycobacterium tuberculosis* fibronectin-binding proteins. *Infect. Immun.* **59**, 2712–2718 (1991).
 36. Robertson, B. D. & Meyer, T. F. Genetic variation in pathogenic bacteria. *Trends Genet.* **8**, 422–427 (1992).
 37. Levitskaya, J. *et al.* Inhibition of antigen processing by the internal repeat region of the Epstein-Barr virus nuclear antigen-1. *Nature* **375**, 685–688 (1995).
 38. Levitskaya, J., Sharipo, A., Leonchiks, A., Ciechanover, A. & Masucci, M. G. Inhibition of ubiquitin/proteasome-dependent protein degradation by the Gly-Ala repeat domain of the Epstein-Barr virus nuclear antigen 1. *Proc. Natl Acad. Sci. USA* **94**, 12616–12621 (1997).
 39. Flynn, J. L., Goldstein, M. A., Treibold, K. J., Koller, B. & Bloom, B. R. Major histocompatibility complex class-II restricted T cells are required for resistance to *Mycobacterium tuberculosis* infection. *Proc. Natl Acad. Sci. USA* **89**, 12013–12017 (1992).
 40. Bloom, B. R. & Fine, P. E. M. in *Tuberculosis: Pathogenesis, Protection, and Control* (ed. Bloom, B. R.) 531–557 (Am. Soc. Microbiol., Washington DC, 1994).
 41. Collins, D. M. In search of tuberculosis virulence genes. *Trends Microbiol.* **4**, 426–430 (1996).
 42. Arruda, S., Bomfim, G., Knights, R., Huima-Byron, T. & Riley, L. W. Cloning of an *M. tuberculosis* DNA fragment associated with entry and survival inside cells. *Science* **261**, 1454–1457 (1993).
 43. Baumler, A. J., Kusters, J. G., Stojkovic, I. & Heffron, F. *Salmonella typhimurium* loci involved in survival within macrophages. *Infect. Immun.* **62**, 1623–1630 (1994).
 44. McMurray, A. A., Sulston, J. E. & Quail, M. A. Short-insert libraries as a method of problem solving in genome sequencing. *Genome Res.* **8**, 562–566 (1998).
 45. Bonfield, J. K., Smith, K. F. & Staden, R. A new DNA sequence assembly program. *Nucleic Acids Res.* **24**, 4992–4999 (1996).
 46. Krogh, A., Mian, I. S. & Haussler, D. A hidden Markov model that finds genes in *E. coli* DNA. *Nucleic Acids Res.* **22**, 4768–4778 (1994).
 47. Baïroch, A., Bucher, P. & Hofmann, K. The PROSITE database, its status in 1997. *Nucleic Acids Res.* **25**, 217–221 (1997).
 48. Altschul, S., Gish, W., Miller, W., Myers, E. & Lipman, D. A basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
 49. Pearson, W. & Lipman, D. Improved tools for biological sequence comparisons. *Proc. Natl Acad. Sci. USA* **85**, 2444–2448 (1988).
 50. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic DNA. *Nucleic Acids Res.* **25**, 955–964 (1997).

Acknowledgements. We thank Y. Av-Gay, F.-C. Bange, A. Danchin, B. Dujon, W. R. Jacobs Jr, L. Jones, M. McNeil, I. Moszer, P. Rice and J. Stephenson for advice, reagents and support. This work was supported by the Wellcome Trust. Additional funding was provided by the Association Française Raoul Follereau, the World Health Organisation and the Institut Pasteur. S.V.G. received a Wellcome Trust travelling research fellowship.

Correspondence and requests for materials should be addressed to B.G.B. (barrell@sanger.ac.uk) or S.T.C. (stcole@pasteur.fr). The complete sequence has been deposited in EMBL/GenBank/DDJB as MTBH37RV, accession number AL123456.

YOURS TO HAVE AND TO HOLD BUT NOT TO COPY

The publication you are reading is protected by copyright law. Photocopying copyright material without permission is no different from stealing a magazine from a newsagent, only it doesn't seem like theft.

If you take photocopies from books, magazines and periodicals at work your employer should be licensed with CLA.

Make sure you are protected by a photocopying licence.



The Copyright Licensing Agency Limited
90 Tottenham Court Road, London W1P 0LP
Telephone: 0171 436 5931 Fax: 0171 436 3986

Nature of the heating mechanism for the diffuse solar corona

E. R. Priest*, C. R. Foley†, J. Heyvaerts‡, T. D. Arber*, J. L. Culhane† & L. W. Acton§

* Mathematical and Computational Sciences Department, St Andrews University, St Andrews KY16 9SS, UK

† Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK

‡ Observatoire, 11 Rue de l'Université, F-6700 Strasbourg, France

§ Physics Department, Montana State University, Bozeman, Montana 59717, USA

The temperature of the Sun's outer atmosphere (the corona) exceeds that of the solar surface by about two orders of magnitude, but the nature of the coronal heating mechanisms has long been a mystery¹. The corona is a magnetically dominated environment, consisting of a variety of plasma structures including X-ray bright points, coronal holes and coronal loops. The latter are closed magnetic structures that occur over a range of scales and are anchored at each end in the solar surface. Large-scale regions of diffuse emission are made up of many long coronal loops². Here we present X-ray observations of the diffuse corona from which we deduce its likely heating mechanism. We find that the observed variation in temperature along a loop is highly sensitive to the spatial distribution of the heating. From a comparison of the observations and models we conclude that uniform heating gives the best fit to the loop temperature distribution, enabling us to eliminate previously suggested mechanisms of low-lying heating near the footpoints of a loop. Our findings favour turbulent breaking and reconnection of magnetic field lines as the heating mechanism of the diffuse solar corona.

Part of the coronal heating question, namely what heats X-ray

bright points, has already been answered because the converging flux model³ has explained their observational properties in a natural way in terms of reconnection of magnetic field lines in the corona. This is driven by solar surface motions of the footpoints of such field lines and accounts for the internal structure of particular bright points⁴. However, the mechanisms for heating coronal loops and coronal holes have not yet been identified⁵.

Observations from the Soft X-ray Telescope on board the Japanese/US/UK Yohkoh satellite have built on previous rocket and Skylab observations and revealed that the corona is a complex magnetohydrodynamic system that is highly time-dependent and three-dimensional², with myriads of magnetic loops continually evolving and interacting. These observations have also given important clues as to how the corona is heated. In particular, a distinction has been revealed between local, time-dependent impulsive components to the heating and global, steadier components on a much larger scale⁶. For example, in active regions there are tiny transient brightenings⁷, which are likely to be driven by reconnection (like X-ray bright points), but they fall short by a factor of 5–10 of being able to heat active-region loops. Furthermore, the hottest loops in active regions (>6 MK) appear to be multiple structures that are interacting by reconnection or tiny cusp-like features⁸ that are closing down by reconnection, as in large eruptive solar flares⁹. The cooler coronal loops (3–5 MK) are much steadier in their emission¹⁰.

Sturrock *et al.*¹¹ studied a large magnetically closed region of the diffuse corona and deduced from filter ratios that the temperature increases with radius from 1.6 MK at the limb to 2.3 MK at 1.5 $R_{\odot}$. They modelled this temperature variation roughly by assuming that the heat flux is constant and the field lines radial. Here we present new observations of a single diffuse region with lower errors and model them in a more realistic way.

We have analysed a particular loop (outlined in Fig. 1a), which can be clearly seen in Fig. 1b. We carefully corrected the Soft X-ray Telescope images for the effect of radiation scattered from other areas of the Sun due to the broad low-level wings of the telescope point-spread function¹². The level of the scattered signal amounts to about 5% of the observed flux at the loop footpoints and 20% at the

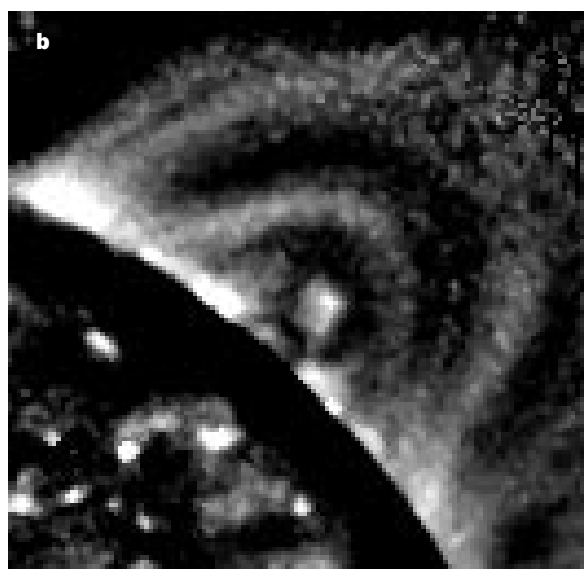
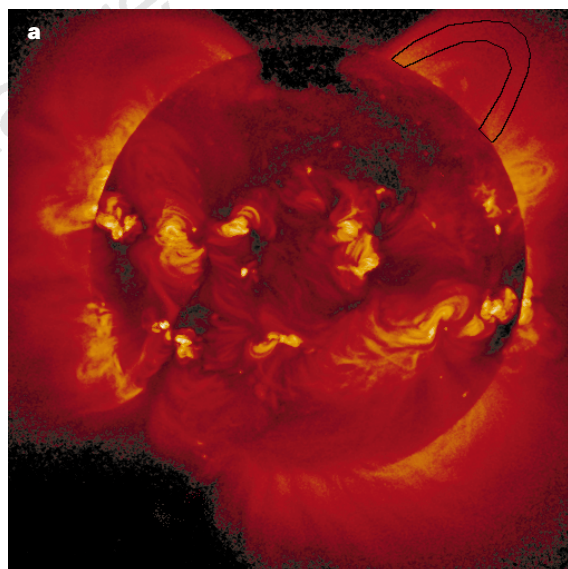


Figure 1 Soft X-ray image of the Sun from the Yohkoh Soft X-ray Telescope on 3 October 1992. **a**, Global image indicating the outer loop of the arcade that we have analysed; **b**, close-up of the region. Yohkoh images such as this reveal diffuse emission from large regions that extend up to 0.8 solar radii above the limb and evolve slowly over long periods of time. Magnetic field extrapolations demonstrate that the largely structureless diffuse emission arises in large-scale

magnetically closed regions and almost always underlies coronal streamers⁶. The visibility of specific loops in an X-ray image depends on the temperature and density contrast along neighbouring magnetic flux tubes. Low-altitude, intensely heated loops are easily distinguished. At large heights, entire loops may be seen (top left and top right), but often only the parts near the base are identifiable, arching round with a convex curvature towards the solar limb (bottom right).

loop top. Temperatures were obtained by the filter-ratio technique. The data were aligned to within 2 arcsec and corrected for the effects of CCD saturation, dark noise and particle hits. For example, Figs 1–3 show the global image and the resulting temperature variation along the length of a loop and with height at the symmetry axis of an arcade. The temperature in such arcades rises rapidly and levels off at typically 2.2–2.3 MK at an altitude of $1.5 R_{\odot}$.

We have compared our observed temperature profiles with a set of standard models for a coronal loop. They incorporate gravity and possess a steady-state energy balance between radiation, thermal conduction along magnetic field lines and a heating that depends on distance along the loop in a variety of different ways. The models start from the summit temperature (a free parameter determined from the observations) and integrate down to 1.5 MK. Over this temperature range we find that the effects of radiation and gravitation on the temperature profile are negligible: radiation here is a factor of about 100 times smaller than conduction in the energy balance, which therefore determines the temperature independently of the pressure and density; gravity makes the pressure and density fall off with height but does not influence the temperature profile. If we continued the model further down and incorporated extra transition-region and chromospheric physics it would make no

difference to the coronal part that we have compared with the observations.

For each parametric form of heating, the most likely parameters and heating values are deduced by a weighted least-squares analysis, by minimizing the sum (χ^2) of the squares of the differences between the observed and model temperatures (normalized with respect to the errors in the observed temperatures). Dividing by the number (n) of degrees of freedom gives a quantity (χ^2/n) that has an expected value of unity if the true model has been identified and an expected value of greater than one otherwise.

The data for the individual loop shown in Fig. 1 have been compared with three different heating models. The first one has a given summit temperature (T_0) and a maximum in its heating rate (H_0) per unit volume at the loop base; it decays exponentially along the loop over a distance $0.1L$, where L is the loop half-length. Minimization with respect to the parameters T_0 and H_0 yielded a very poor fit for the temperature profile (Fig. 2a) with a χ^2/n of 5.4 and a significance level less than 0.1%. The second model has a heat source concentrated at the loop summit, and did not give a good fit (Fig. 2b): the value of χ^2/n for this fitted model still exceeds unity (1.7), although the evidence against it is not strong (significance level of 10%). The third model has uniform heating (H_0), and the resulting minimization gave a rather good fit with a χ^2/n of 0.89 and a significance level of 51% (Fig. 2c). This is strong evidence against heating concentrated near the loop base and suggests that heating uniformly distributed along the loop^{13,14} is more likely than heating concentrated at the summit.

We next modelled the arcade as a potential magnetic field with a series of loops that are arcs of circles, so that the magnetic field falls

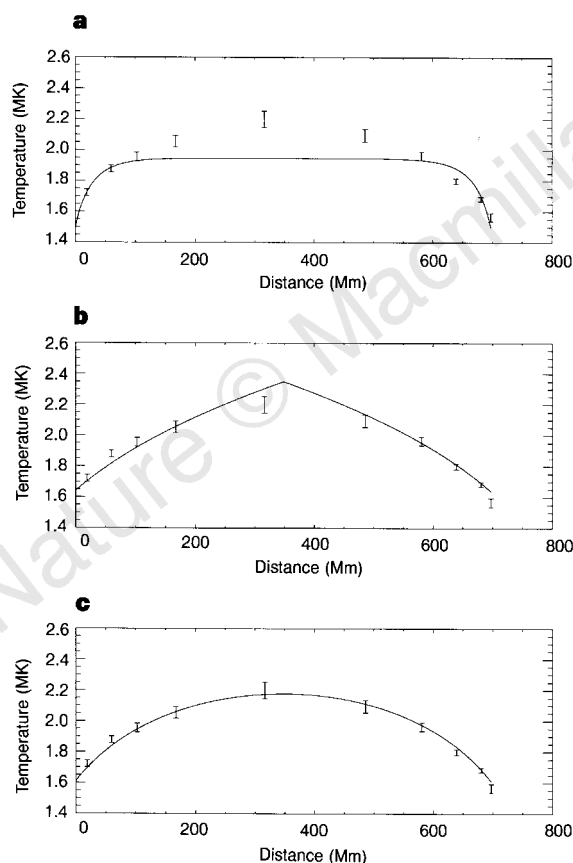


Figure 2 Observed temperature (in MK = 10^6 K) as a function of distance (in Mm = 1,000 km) along the length of the loop from one coronal footpoint to the other with the error as shown. Also included are the best-fit model temperatures for **a**, heating that decays from the footpoints exponentially over a distance of $0.1L$, where L is the loop half-length; **b**, heating concentrated at the summit; and **c**, uniform heating. The errors in the temperature reflect the statistical uncertainty in the data and do not include systematic uncertainty due to pre-launch filter calibration (which is expected to be less than 2%). In any case, it is the profile of temperature that matters for the present analysis rather than the absolute values, because a global increase or decrease of temperature would not change the shape of the profile. The errors are smaller than previous ones¹¹ because we have sampled typically 10 times more pixels and averaged over them.

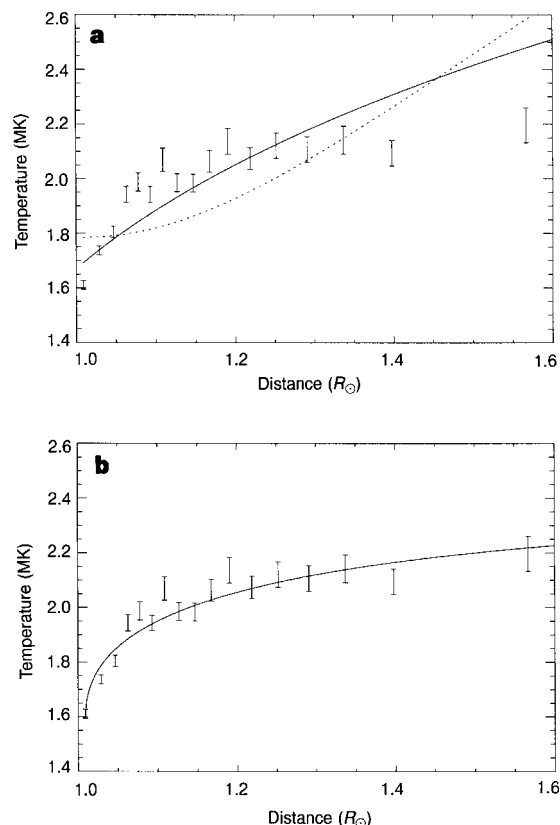


Figure 3 Observed temperature as a function of height at the summits of the loops that comprise the arcade whose outermost loop is shown in Fig. 1. The temperature in MK is shown as a function of distance from the solar centre (in units of $1 R_{\odot} = 700$ Mm) for models with a flux that is **a**, proportional to loop length (dotted) or uniform (solid) and **b**, proportional to the square of the magnetic field of an arcade model.

off inversely with distance from an axis below the solar surface. Each of the loops is heated uniformly and viewed face-on with no assumed variation in the line of sight, so that the temperature at a given altitude on the axis of the arcade is the value at the summit of the particular loop that reaches that altitude. First, we assumed the energy is deposited uniformly per unit length, so that the heating flux is proportional to loop length, and found a very poor fit with a χ^2/n of 4.95 and a significance level below 0.1% (dotted curve in Fig. 3a). Next, we assumed instead the same heating flux on each loop and found a better (but still poor) fit with a χ^2/n of 1.83 and a significance level of 2.5% (solid curve in Fig. 3a). Finally, we considered a heating that is proportional to the square of the magnetic field strength and so decreases with altitude. This gave an excellent fit with a χ^2/n of 0.49 and a significance level of 94% (Fig. 3b).

Several different mechanisms for heating the solar corona have previously been proposed and they produce different forms of heating, so that the above result about the heating probably being distributed uniformly along large-scale diffuse loops has important implications about which mechanism is likely. In particular, we may deduce that the diffuse corona is heated locally *in situ* and rule out the previously reasonable suggestion that it is just a response to low-lying heating near the loop feet, either by the interaction of small loops that takes place in bright points or by waves of very small wavelength dumping their energy low down in the atmosphere.

Long-wavelength magnetic waves were another prime candidate for heating large coronal loops^{15,16}, driven resonantly by the observed solar surface motions. It has been suggested that they may dissipate and dump their energy rather non-uniformly in the coronal plasma by two mechanisms, known as phase mixing^{17,18} and resonant absorption^{19–21}. The fundamental standing waves (which carry the most energy) have nodes near the loop feet and their greatest amplitude near the loop summits, where the heating is therefore concentrated non-uniformly, and so such waves are unlikely to explain our observations. However, although waves seem unlikely, they cannot definitely be ruled in or out using the present technique until the spatial distribution of heating by wave mechanisms has been carefully calculated.

At present, the evidence is more in favour of another mechanism that does tend to heat the plasma uniformly, namely stochastic or turbulent dissipation in many small current sheets^{22–24}. These may form in response to either the slow braiding of the magnetic footpoints in the solar surface^{25,26} or local small-scale instabilities that are driven by such surface motions. In either case a state of magnetic turbulence is set up²⁷, which has recently been modelled in three-dimensional numerical experiments^{28–30}.

We have here proposed a new two-part approach to trying to solve the coronal heating problem, namely first of all to use observed temperature profiles to deduce the form of the heating in coronal loops and arcades and secondly to use that heating form to deduce the likely heating mechanism. The application of this approach here suggests that large-scale diffuse coronal loops are much more likely to be heated rather uniformly by turbulent magnetic reconnection than by preferential heating at the loop feet or loop summit. In future, it will be interesting to see whether or not the same is true for other coronal structures. In addition, it is hoped that this will act as an incentive to proponents of particular mechanisms to deduce in detail the resulting form of heating produced by their models and to observers to refine their diagnostics so as to deduce the temperature profiles with as low an error as possible. □

Received 16 May 1997; accepted 17 March 1998.

1. Ulmschneider, P., Priest, E. R. & Rosner, R. *Mechanisms of Chromospheric and Coronal Heating* (Springer, Berlin, 1991).
2. Culhane, J. L. Yohkoh observations of the solar corona. *Adv. Space Res.* **19**, 1839–1848 (1997).
3. Priest, E. R., Parnell, C. E. & Martin, S. F. A converging flux model of an x-ray bright point and an associated cancelling magnetic feature. *Astrophys. J.* **427**, 459–474 (1994).
4. Parnell, C. E., Golub, L. & Priest, E. R. The 3D structure of x-ray bright points. *Solar Phys.* **151**, 57–74 (1994).

5. Jordan, C. Modelling of solar coronal loops. *Mem. Soc. Astron. Ital.* **63**, 605–620 (1992).
6. Acton, L. W. Comparison of Yohkoh x-ray and other solar activity parameters for November 1991 to November 1995. *Astron. Soc. Pacific Conf. Ser.* **109**, 45–54 (1996).
7. Shimizu, T., Tsuneta, S., Acton, L. W., Lemen, J. R. & Uchida, Y. Transient brightenings in active regions observed by the Yohkoh soft X-ray telescope. *Publ. Astron. Soc. Japan* **44**, L147–L153 (1992).
8. Yoshida, T. & Tsuneta, S. Temperature structure of solar active regions. *Astrophys. J.* **459**, 342–346 (1996).
9. Tsuneta, S. & Lemen, J. R. in *Physics of Solar and Stellar Coronae* 113–130 (eds Linsky, J. & Serio, S.) (Kluwer, Dordrecht, 1993).
10. Kano, R. & Tsuneta, S. Temperature distributions and energy scaling law of solar coronal loops obtained with Yohkoh. *Publ. Astron. Soc. Japan* **48**, 535–543 (1996).
11. Sturrock, P. A., Wheatland, M. S. & Acton, L. W. Yohkoh soft x-ray telescope images of the diffuse solar corona. *Astrophys. J.* **461**, L115–L117 (1996).
12. Foley, C. A., Culhane, J. L., Weston, D., Acton, L. W. & Hara, H. *Solar Phys.* (submitted).
13. Neupert, W., Nakagawa, Y. & Rust, D. M. Energy balance in a magnetically confined coronal structure observed by OSO-7. *Solar Phys.* **43**, 359–376 (1975).
14. Landini, M. & Monsignori-Fossi, B. C. A loop model of active coronal regions. *Astron. Astrophys.* **42**, 213–220 (1975).
15. Hollweg, J. V. in *Solar Wind 5* (NASA CP 2280) 5–21 (NASA, Washington DC, 1983).
16. Roberts, B. in *Hydromagnetics of Sun* (ESA SP-220) 137–145 (ESA, Noordwijk, 1984).
17. Heyvaerts, J. & Priest, E. R. Coronal heating by phase-mixed shear Alfvén waves. *Astron. Astrophys.* **117**, 220–234 (1983).
18. Hood, A. W., Ireland, J. & Priest, E. R. Heating of coronal holes by phase mixing. *Astron. Astrophys.* **318**, 957–961 (1997).
19. Goedbloed, J. P. *Lecture Notes on Ideal MHD* (Rijnhuizen Report) 83–145 (1983).
20. Goossens, M. in *Advances in Solar System MHD* (eds Priest, E. R., Hood, A. W.) 137–172 (Cambridge Univ. Press, 1991).
21. Davila, J. M. Heating of the solar corona by resonant absorption of Alfvén waves. *Astrophys. J.* **317**, 514–521 (1987).
22. Tucker, W. H. Heating of solar active regions by magnetic energy dissipation: the steady state case. *Astrophys. J.* **186**, 285–289 (1975).
23. Levine, R. Evidence for opposed currents in active-region loops. *Solar Phys.* **46**, 159–170 (1976).
24. Galeev, A. A., Rosner, R., Serio, S. & Vaiana, G. S. Dynamics of coronal structures: magnetic field related heating and loop energy balance. *Astrophys. J.* **243**, 301–308 (1981).
25. Parker, E. N. *Spontaneous Current Sheets in Magnetic Fields* (Oxford Univ. Press, 1994).
26. Berger, M. A. in *Advances in Solar System MHD* (eds Priest, E. R. & Hood, A. W.) 241–256 (Cambridge Univ. Press, 1991).
27. Heyvaerts, J. & Priest, E. R. A self-consistent turbulent model for coronal heating. *Astrophys. J.* **390**, 297–308 (1993).
28. Einaudi, G., Velli, M., Politano, H. & Pouquet, A. Energy release in a turbulent corona. *Astrophys. J.* **457**, L113–L116 (1996).
29. Hendrix, D. L., Van Hoven, G., Mikic, Z. & Schnack, D. D. The viability of Ohmic dissipation as a coronal heating source. *Astrophys. J.* **470**, 1192–1197 (1996).
30. Galsgaard, K. & Nordlund, A. Heating and activity of the solar corona 1. Boundary shearing of an initially homogeneous magnetic field. *J. Geophys. Res.* **101**, 13445–13460 (1996).

Acknowledgements. We are most grateful for financial support from the UK Particle Physics and Astronomy Research Council and to S. T. Buckland for helpful suggestions. Part of the work was carried out while one of us (J.L.C.) was a visiting professor at the Institute for Space and Astronautical Science (ISAS) of Japan. C.R.F. was supported by a Research Studentship from PPARC. NASA supported the work of L.W.A. The Yohkoh mission and its continued operation are projects of ISAS in Japan.

Correspondence and requests for materials should be addressed to E.R.P. (e-mail: eric@dcs.st-and.ac.uk).

Chemical processing in the coma as the source of cometary HNC

William M. Irvine*, Edwin A. Bergin†, James E. Dickens*, David Jewitt‡, Amy J. Lovell*, Henry E. Matthews§, F. Peter Schloerb* & Matthew Senay*

* Five College Radio Astronomy Observatory, 619 LGRC, University of Massachusetts, Amherst, Massachusetts 01003, USA

† Smithsonian Astrophysical Observatory, MS-66, 60 Garden Street, Cambridge, Massachusetts 02138-1596, USA

‡ Institute for Astronomy, University of Hawaii, 2680 Woodlawn Drive, Honolulu, Hawaii 96822, USA

§ Herzberg Institute of Astrophysics, National Research Council of Canada, Ottawa, Ontario K1A 0R6, Canada

|| Joint Astronomy Centre, 660 N. A'ohoku Place, University Park, Hilo, Hawaii 96720, USA

The discovery of hydrogen isocyanide (HNC) in comet Hyakutake with an abundance (relative to hydrogen cyanide, HCN) similar to that seen in dense interstellar clouds raised the possibility that these molecules might be surviving interstellar material¹. The preservation of material from the Sun's parent molecular cloud would provide important constraints on the processes that took place in the protostellar nebula. But another possibility is that

HNC is produced by photochemical processes in the coma, which means that its abundance could not be used as a direct constraint on conditions in the early Solar System. Here we show that the HNC/HCN ratio determined for comet Hale–Bopp varied with heliocentric distance in a way that matches the predictions of models of gas-phase chemical production of HNC in the coma, but cannot be explained if the HNC molecules were coming from the comet's nucleus. We conclude that HNC forms mainly by chemical reactions in the coma, and that such reactions need to be considered when attempting to deduce the composition of the nucleus from observations of the coma.

The $J = 4-3$ pure rotational transitions^{2,3} of both HCN and HNC were observed in comet Hale–Bopp (C/1995 O1) at the James Clerk Maxwell Telescope (Fig. 1). The lines of both species were either observed simultaneously or within, at most, one hour of each other at approximately equal zenith angles. Such timing is very important, as intra-day variations in at least the HCN emission arise as the comet rotates and undergoes outbursts^{4,5}.

Table 1 gives the observed intensities integrated over the line frequencies, $I(\text{HCN})$ and $I(\text{HNC})$. The ratio, $R_{\text{thin}} = I(\text{HNC})/I(\text{HCN})$, will closely approximate the ratio of column densities (average number of molecules per cm^2 in the telescope beam), apart from effects of optical depth (Table 1). In fact, R_{thin} will be an upper limit to the true ratio of column densities, as the HCN lines are sufficiently strong to ensure that we need to consider their opacity. It is clear that R_{thin} increases strongly as the comet approaches the Sun. Such an increase is also apparent in the results published by Biver *et al.*^{6,7}, who, however, do not discuss opacity effects.

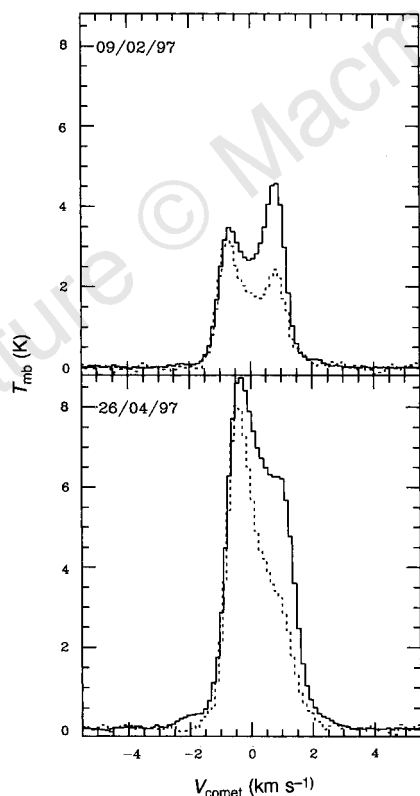


Figure 1 The $J(4-3)$ lines of HNC (dashed lines) and HCN (solid lines, spectra multiplied by 0.25) for comet Hale–Bopp. Observations were made with the James Clerk Maxwell Telescope on the indicated dates. The velocity scale is with respect to the nucleus, and the intensity scale is main beam brightness temperature. The double-peaked profiles that are often observed are indicative of an expanding and resolved shell of gas, such as those seen around many evolved stars²⁴.

A critical question is whether the heliocentric variation in R_{thin} might be solely an effect of the increasing opacity of the HCN line, as the comet approaches the Sun and outgassing increases. Determining the optical depth of the HCN emission is difficult, as it involves assumptions about the distribution and kinematics of the coma gas, and the excitation of HCN. There are, in principle, two mitigating factors in such determinations: the $J = 4-3$ transition of HCN has two weak hyperfine components (each $\sim 2\%$ of the total line strength) that are partially resolved from the three strongest, overlapping hyperfine components; also, we have some observations of the rarer isotopomer, H^{13}CN .

An estimate of the possible saturation was obtained by assuming that the observed HCN emission originated in a homogeneous slab of gas, so that the excitation temperature (T_x) was constant in the observed region and the line width was the result of random mass motions on a scale smaller than the beam of our telescope. In the present case, this approach should generally overestimate possible opacity effects, because in a more realistic expanding coma the systematic velocity gradients will make it easier for photons emitted in the interior of the coma to escape. The results show that the HCN line has at most a moderate opacity, with maximum values $1 < \tau < 2$, consistent with the clearly optically thin profiles for the $J = 1-0$ transition⁸ and with the relative intensities of the HCN and H^{13}CN lines when both were observed. A sample fit is shown in Fig. 2.

Correcting the optically thin abundance ratio for partial saturation of the HCN lines, and assuming that the HNC lines are optically thin, produces an approximate lower bound on the column density ratio: $R_{\text{thick}} = a(\tau)R_{\text{thin}}$, where $a(\tau) = (1 - e^{-\tau})/\tau$, τ is the optical depth for the HCN line, and we have adopted those fit parameters that produce the largest values of τ for observations on a given date.

The fundamental question we wish to resolve is whether HCN and HNC are present primarily as ices in the cometary nucleus, rather than being produced in the coma. Laboratory experiments show that, once water ice begins to sublime rapidly (heliocentric distance $r \leq 4$ AU), the sublimation of trace nuclear constituents is controlled by that of H_2O (ref. 9); this is consistent with the constant relative production rates of HCN and H_2O as a function of r

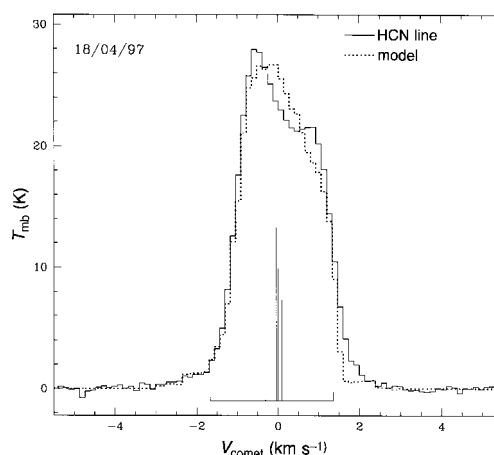


Figure 2 Comparison of an observed $J = 4-3$ HCN line (solid line; for indicated date) with a radiative transfer model (dashed line). Here we used the $J = 1-0$, $F = 2-1$, line profile to estimate the optical depth of the $J = 4-3$ line. We note hyperfine components of $J = 4-3$ line at positions indicated by vertical markers; height of markers shows relative line strength. We solved iteratively for the values of linewidth ΔV and line centre optical depth (τ) which best matched the total observed HCN line shape, including the weak hyperfine components. In our solutions the deduced values of τ were not very dependent on the normalized line profile or on whether or not the line excitation temperature was fixed or was a free parameter (a more complete analysis will be given elsewhere)¹³.

Table 1 Observations and abundance ratios for HCN and HNC

Date (dd/mm/yy)	r (AU)	Δ (AU)	$I(\text{HCN})$ (K km s ⁻¹)	$I(\text{HNC})$ (K km s ⁻¹)	τ_{max}	R_{thin} (%)	R_{thick} (%)
14/07/96	3.779	2.798	0.79 (0.02)	<0.09	—	<11.4	—
06/09/96	3.177	2.850	1.65 (0.02)	<0.02	—	<1.2	—
30/11/96	2.157	2.933	3.23 (0.03)	0.22 (0.02)	0.19 (0.09)	6.8	6.2
18/01/97	1.539	2.273	10.78 (0.03)	1.48 (0.05)	0.30 (0.12)	13.7	11.9
19/01/97	1.527	2.255	10.96 (0.06)	1.44 (0.09)	0.31 (0.11)	13.1	11.3
08/02/97	1.286	1.866	21.93 (0.05)	3.82 (0.06)	0.52 (0.13)	17.4	13.6
09/02/97	1.274	1.846	21.12 (0.02)	3.55 (0.02)	0.50 (0.11)	16.8	13.2
09/02/97*	1.274	1.846	7.62 (0.02)	1.45 (0.03)	0.16 (0.05)	19.0	17.6
16/02/97	1.196	1.711	28.52 (0.03)	4.72 (0.02)	1.22 (0.13)	16.6	9.6
09/03/97	1.003	1.387	31.91 (0.02)	6.46 (0.03)	—	20.2	—
15/03/97	0.964	1.338	49.75 (0.12)	11.01 (0.16)	0.95 (0.21)	22.1	14.3
22/03/97	0.932	1.315	79.20 (0.11)	20.96 (0.18)	1.81 (0.31)	26.5	12.2
18/04/97	0.962	1.555	37.30 (0.07)	8.30 (0.08)	0.72 (0.10)	22.3	15.9
26/04/97	1.015	1.685	43.77 (0.03)	7.52 (0.02)	0.91 (0.13)	17.2	11.3
08/06/97	1.473	2.369	22.17 (0.05)	3.06 (0.05)	—	13.8	—
20/06/97	1.623	2.517	9.64 (0.02)	0.78 (0.02)	0.53 (0.11)	8.1	6.3
14/07/97	1.927	2.747	6.85 (0.03)	0.59 (0.02)	0.13 (0.03)	8.6	8.0
29/08/97	2.500	2.996	2.88 (0.03)	<0.05	—	<1.7	—
30/08/97	2.512	2.999	2.52 (0.02)	<0.05	—	<2.0	—

Date (dd/mm/yy)	r (AU)	Δ (AU)	$I(\text{H}^{13}\text{CN})$ (K km s ⁻¹)
09/02/97	1.274	1.846	0.29 (0.02)
16/02/97	1.196	1.711	0.55 (0.01)
22/03/97	0.932	1.315	1.21 (0.03)
26/04/97	1.015	1.685	0.69 (0.02)

Shown are the integrated intensities (I) for the $J = 4-3$ transitions of HCN and HNC in comet Hale-Bopp, uncorrected for beam efficiency, as a function of heliocentric distance r and geocentric distance Δ ; τ_{max} is the maximum value of line-centre optical depth for the HCN line found from fitting the hyperfine structure; R_{thin} and R_{thick} are upper and lower limits, respectively, to the HNC/HCN column density ratios (see text). Detailed observing and calculation procedures will be described elsewhere¹³. In deriving abundances from the observed spectra, we must consider the relative excitation of the $J = 4$ levels of HCN and HNC and the possibility that the stronger HCN line may be partially saturated. With our 14-arcsec (FWHP, or full width half power) beam (corresponding to 10,000 km at the comet) we are observing the inner coma, where the excitation of both HCN and HNC should be dominated by collisions rather than by radiative excitation²⁰, since the collision-dominated region of the coma is probably a few times 10,000 km in radius²¹ even at $r \approx 2$ AU for such an active comet (the gas production rate at that distance⁶ was $\sim 10^{30}$ s⁻¹). Then, because the relevant molecular parameters (rotational constant, and hence transition frequency, and electric dipole moment) of HNC and HCN are very nearly equal, their excitation and hence their population distribution over accessible energy levels will be nearly equal, as in dense interstellar clouds^{22,23}. In consequence, the ratio of integrated line intensities $I(\text{HNC})/I(\text{HCN})$ will closely approximate the ratio of column densities, apart from effects of optical depth.

* Offset from the nucleus by right ascension 10°, declination 10°.

observed for comet Hale-Bopp⁶. Thus, if HNC and HCN were primarily parent molecules present in the nucleus, their production into the gas phase would vary together, so that the abundance ratio R would not depend on r .

The limits that we have obtained for the variation of R with r are shown in Fig. 3 and listed in Table 1. Even when the varying optical depth of the HCN line is taken into account, the HNC/HCN ratio clearly increases with decreasing heliocentric distance. We conclude that HNC is primarily produced by processes dependent on the strength of the solar radiation field, including the gas production rate. Moreover, the data for $r > 2$ AU, when both the HCN and HNC lines should be optically thin, indicate that any intrinsic HNC/HCN ratio in the nuclear ices of comet Hale-Bopp is less than ~ 0.02 (3σ).

Among the processes which might produce HNC as the comet approaches the Sun, the most plausible are photodissociation of a heavier ('parent') molecule, evaporation or sputtering from organic ('CHON') grains in the coma, and gas-phase chemistry in the coma¹.

We have investigated the possibility that HNC is the 'daughter' of an unknown parent by calculating the distribution in the coma of HNC for various assumed parent lifetimes against photodissociation, using a Monte Carlo model to track the location of the HNC molecules¹⁰. For comet Hale-Bopp, maps of the coma distribution of HCN emission are consistent with a nuclear origin^{4,5,8}. If the HCN is subliming directly from the nucleus while the HNC is produced in the coma, their distributions as a function of coma position will differ; the abundance ratio must be obtained by convolving the calculated distribution with the antenna beam for the appropriate date. The calculated results (Fig. 3) show a dependence of the HNC/HCN ratio on heliocentric distance which does not match the observations well, irrespective of the parent lifetime.

In contrast, gas-phase chemistry must be important for a comet as active as Hale-Bopp. Certainly HCO^+ , emission from which was mapped for the first time in comet Hale-Bopp^{4,11}, must be produced in this way. We have investigated this situation with a time-

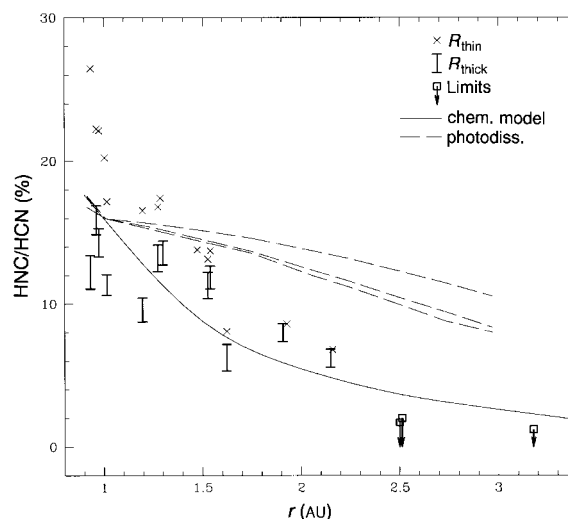


Figure 3 Dependence of the HNC/HCN ratio on heliocentric distance r . R_{thin} is the observed column density ratio assuming that both lines are optically thin, and R_{thick} is the corresponding ratio with the maximum estimate of the HCN opacity correction (error bars show $\pm 1\sigma$ about this value). Open squares are 3σ upper limits. Dashed lines give the calculated¹⁰ HNC/HCN ratio for the case that HCN is subliming from the nucleus while HNC is formed by photodissociation of an unknown nuclear parent which is sufficiently abundant to produce the same value of the ratio as our chemical model at 1 AU. Upper, middle and lower dashed curves are for parent lifetimes in the solar radiation field equal to 0.1, 1.0 and 10 times that of HCN (ref. 25), respectively. The results of our chemical model are shown as a solid line. The model water production rate (molecules s⁻¹) was set equal to $10^{31} r^{-1.6}$ (here r is in AU; this value is the approximate rate given in ref. 26); the abundances in the nuclear ices were set to the reported relative production rates at 1.5 AU (that is, when the water ice is subliming rapidly^{6,11}), except that the nuclear abundance of HNC was set equal to zero; and the outflow velocity was taken to be 1 km s⁻¹.

dependent interstellar chemical model¹², adapted to the cometary conditions (details will be published elsewhere¹³; see also Fig. 3 legend). We find that HCN undergoes proton transfer from H_3O^+ to produce HCNH^+ , which can in turn react by proton transfer with H_2CO and NH_3 , or by dissociative recombination with electrons to give HNC (and HCN). Dissociative recombination is the most important pathway according to our calculations. Although there are a number of somewhat uncertain parameters in the model, the similar production and destruction pathways for HNC and HCN make the calculated HNC/HCN ratio rather insensitive to most of these parameters. For example, we find that varying the initial (nuclear) abundances of various nitrogen-containing species over a wide range ($0 < \text{N}_2 < 0.3$; $0.002 < \text{NH}_3 < 0.05$, relative to H_2O) has little effect on the HNC/HCN ratio.

The parameters crucial to the calculated HNC/HCN ratio are the branching ratios for the principal HNC production route, $\text{HCNH}^+ + e$. There are no laboratory measurements of these ratios, and the standard values used for interstellar chemistry assume that this dissociative recombination yields $\text{HCN}:\text{HNC}:\text{CN}$ in the ratios 0.25:0.25:0.5 (ref. 14). However, very recent quantum-chemical calculations suggest that the production of HNC should exceed that of HCN and that little CN will be produced¹⁵. This is consistent with the conclusions from extensive observations of HCN and HNC isotopomers in the interstellar medium, from which the branching ratios are deduced to be¹⁶ $\text{HCN}:\text{HNC}:\text{CN} = 0.4:0.6:0$. We adopt here the slightly more conservative ratios $\text{HCN}:\text{HNC}:\text{CN} = 0.45:0.55:0$.

The results of this chemical model are plotted in Fig. 3. The model matches the observations very well, both in terms of the maximum value of the HNC/HCN ratio and in terms of the dependence of this ratio on heliocentric distance r . The nonlinear character of the dependence on r seems to be a result of the multi-stage chemical process which produces HNC and the density-squared dependence of the bimolecular reactions involved.

A source for HNC from the CHON grains^{17,18} in the coma might involve multiple processes which could conceivably mimic the observed heliocentric dependence, and can perhaps not be completely ruled out. However, we feel that HNC in Hale-Bopp must be produced in significant measure by gas-phase, ion-molecule chemistry initiated by the photoionization of H_2O . A similar conclusion has been reached with an independent chemical model by Rodgers and Charnley¹⁹, and our chemical model also predicts an abundance and coma distribution of HCO^+ in agreement with the observations¹¹. We conclude that HNC is, to our knowledge, the first neutral cometary molecule whose existence in active comets can be ascribed in large part to chemical processes in the coma. As a corollary, it follows that in such comets the effects of chemistry in the coma must be included in deducing the original composition of the nucleus. □

Received 9 October 1997; accepted 30 March 1998.

- Irvine, W. M. *et al.* Spectroscopic evidence for interstellar ices in comet Hyakutake. *Nature* **383**, 418–420 (1996).
- Lovas, F. J. Recommended rest frequencies for observed interstellar molecular microwave transitions—1991 revision. *J. Phys. Chem. Ref. Data* **21**, 181–271 (1992).
- Bockelée-Morvan, D., Padman, R., Davies, J. K. & Crovisier, J. Observations of submillimeter lines of CH_3OH , HCN, and H_2CO in comet P/Swift-Tuttle with the James Clerk Maxwell Telescope. *Planet. Space Sci.* **42**, 655–662 (1994).
- Lovell, A. J. *et al.* HCO⁺ imaging of comet C/1995 O1 Hale-Bopp. *Astrophys. J. Lett.* **497**, L117–L121 (1998).
- Wink, J. *et al.* Evidence for extended sources and temporal modulations in molecular observations of C/1995 O1 (Hale-Bopp) at the IRAM interferometer. *Earth Moon Planets Abstr.* (in the press).
- Biver, N. *et al.* Evolution of the outgassing of Comet Hale-Bopp (C/1995 O1) from radio observations. *Science* **275**, 1915–1918 (1997).
- Biver, N. *et al.* Long-term evolution of the outgassing of comet Hale-Bopp from radio observations. *Earth Moon Planets* (in the press).
- Irvine, W. M. *et al.* In *The Far Infrared and Submillimetre Universe* (ed. Wilson, A.) 277–280 (SP-401, ESA, Noordwijk Netherlands, 1997).
- Notesco, G. & Bar-Nun, A. Trapping of methanol, hydrogen cyanide, and *n*-hexane in water ice, above its transformation temperature to the crystalline form. *Icarus* **126**, 336–341 (1997).
- Tacconi-Garman, L. E., Schloerb, F. P. & Claussen, M. J. High spectral resolution observations and kinematic modeling of the 1667 MHz hyperfine transition of OH in comets Halley (1982i), Giacobini-Zinner (1984e), Hartley-Good (1985i), Thiele (1985m), and Wilson (1986i). *Astrophys. J.* **364**, 672–686 (1990).

- Lovell, A. J. *et al.* HCO⁺ ion-molecule chemistry in comet C/1995 O1 Hale-Bopp. *Earth Moon Planets* (in the press).
- Bergin, E. A. & Langer, W. D. Chemical evolution in pre-protostellar and protostellar cores. *Astrophys. J.* **486**, 316–328 (1997).
- Irvine, W. H. *et al.* Chemistry in cometary comae. *Faraday Discuss.* **109** (in the press).
- Herbst, E. What are the products of polyatomic ion-electron dissociative recombination reactions? *Astrophys. J.* **222**, 508–516 (1978).
- Shiba, Y., Hirano, T., Nagashima, U. & Ishii, K. Potential energy surfaces and branching ratio of the dissociative recombination reaction $\text{HCNH}^+ + e^-$: An *ab initio* molecular orbital study. *J. Chem. Phys.* **108**, 698–705 (1998).
- Hirota, T., Yamamoto, S., Mikami, H. & Ohishi, M. Abundances of HCN and HNC in dark cloud cores. *Astrophys. J.* (in the press).
- Mumma, M. J., Weissman, P. R. & Stern, S. A. in *Protostars and Planets III* (eds Levy, E. H. & Lunine, J. I.) 1177–1252 (Univ. Arizona Press, Tucson, 1993).
- Fomenkova, M. in *From Stardust to Planetesimals* (eds Pendleton, Y. J. & Tielens, A. G. G. M.) 415–421 (ASP Conf. Ser. 122, Astron. Soc. Pacific, San Francisco, 1997).
- Rodgers, S. D. & Charnley, S. B. HNC and HCN in comets. *Astrophys. J. Lett.* (in the press).
- Crovisier, J. Rotational and vibrational synthetic spectra of linear parent molecules in comets. *Astron. Astrophys. Suppl.* **68**, 223–258 (1987).
- Festou, M. C., Rickman, H. & West, R. M. Comets II. Models, evolution, origin and outlook. *Astron. Astrophys. Rev.* **5**, 37–163 (1993).
- Goldsmith, P. F., Langer, W. D., Ellder, J., Irvine, W. M. & Kollberg, E. Determination of the HNC to HCN abundance ratio in giant molecular clouds. *Astrophys. J.* **249**, 524–531 (1981).
- Goldsmith, P. F., Irvine, W. M., Hjalmarson, Å. & Ellder, J. Variations in the HCN/HNC abundance ratio in the Orion molecular cloud. *Astrophys. J.* **310**, 383–391 (1986).
- Olofsson, H., Johansson, L. E. B., Hjalmarson, Å. & Nguyen-Quang-Rieu. High sensitivity molecular line observations of IRC +10216. *Astron. Astrophys.* **107**, 128–144 (1982).
- Schmidt, H. U., Wegman, R., Huebner, W. F. & Boice, D. C. Cometary gas and plasma flow with detailed chemistry. *Comput. Phys. Commun.* **49**, 17–59 (1988).
- Proc. 1st Int. Conf. on Comet Hale-Bopp* (to be published as a special issue of *Earth, Moon, Planets*).

Acknowledgements. We thank B. Marsden and D. Tholen for rapidly providing us with the best available cometary elements, and the staff at the JCMT for their assistance with the observations. This work was partly supported by the NSF and NASA.

Correspondence and requests for materials should be addressed to W.M.I. (e-mail: irvine@fcr1.phast.umass.edu).

Electronic liquid-crystal phases of a doped Mott insulator

S. A. Kivelson*, E. Fradkin† & V. J. Emery‡

* Department of Physics, University of California Los Angeles, Los Angeles, California 90095, USA

† Department of Physics, University of Illinois, Urbana, Illinois 61801-3080, USA

‡ Brookhaven National Laboratory, Upton, New York 11973-5000, USA

The character of the ground state of an antiferromagnetic insulator is fundamentally altered following addition of even a small amount of charge¹. The added charge is concentrated into domain walls across which a π phase shift in the spin correlations of the host material is induced. In two dimensions, these domain walls are 'stripes' which can be insulating^{2,3} or conducting^{4–6}—that is, metallic 'rivers' with their own low-energy degrees of freedom. However, in arrays of one-dimensional metals, which occur in materials such as organic conductors⁷, interactions between stripes typically drive a transition to an insulating ordered charge-density-wave (CDW) state at low temperatures. Here it is shown that such a transition is eliminated if the zero-point energy of transverse stripe fluctuations is sufficiently large compared to the CDW coupling between stripes. As a consequence, there should exist electronic quantum liquid-crystal phases, which constitute new states of matter, and which can be either high-temperature superconductors or two-dimensional anisotropic 'metallic' non-Fermi liquids. Neutron scattering and other experiments in the copper oxide superconductor $\text{La}_{1-x}\text{Nd}_x\text{Sr}_x\text{CuO}_4$ already provide evidence for the existence of these phases in at least one class of materials.

Classical liquid crystals⁸ are phases that are intermediate between a liquid and a solid, and spontaneously break the rotation and/or translation symmetries of free space. The proposed electronic liquid crystals are quantum analogues of these phases in which the ground state is intermediate between a liquid, where quantum fluctuations

are large, and a crystal, where they are small. Because the electrons exist in a solid, it is the point-group symmetry of the host crystal that is spontaneously broken, rather than the symmetry of free space. The discussion here will be restricted to a two-dimensional (or quasi-two-dimensional) square lattice, but the generalization to other crystalline symmetries is straightforward. The various zero-temperature phases can then be classified as follows. (1) The liquid phase breaks no spatial symmetries and, in the absence of disorder, is a conductor or a superconductor. (2) The nematic phase breaks the four-fold rotation symmetry of the lattice, but leaves both translation and reflection symmetries unbroken; it is an anisotropic liquid with an axis of orientation. (3) The smectic phase breaks translational symmetry in only one (symmetry) direction. Along the other direction, it has the character of an electron liquid; the smectic order is necessarily commensurate with the underlying lattice at zero temperature. (4) The crystalline phase breaks translation symmetry and has the character of an electronic 'solid'; that is, it is insulating.

We now show how these phases arise, assess their stability, and explore some consequences; at the end we discuss their experimental 'signatures'. We first consider an isolated metallic stripe in a Mott insulator. Such a river of charge is a prototypical example of the one-dimensional electron gas in an active environment⁹, about which much is known. This system is generally 'quantum critical': that is, as the temperature $T \rightarrow 0$, the correlation length diverges and correlation functions fall off as a power of the distance. Frequently there is a gap Δ_s in the spin excitation spectrum, so the only low-energy degrees of freedom are the CDW and the dual superconducting fluctuations whose susceptibilities diverge as $T \rightarrow 0$, as:

$$\chi_{\text{CDW}} \approx \Delta_s T^{-(2-K_c)}; \chi_{\text{SC}} \sim \Delta_s T^{-(2-1/K_c)} \quad (1)$$

where K_c is a non-universal critical exponent¹⁰ which depends on the magnitude and sign of the interactions and satisfies $0 < K_c < 1$ for repulsive interactions.

Direct evidence¹¹⁻¹⁶ of stripe correlations has been obtained over the past few years from neutron-scattering experiments on the copper oxide superconductors, which are the best-studied examples of doped antiferromagnets. It is thus reasonable to ask whether these stripes are relevant for the mechanism of high-temperature superconductivity⁹. Typically, in conventional superconductors, the superconducting gap and transition temperature, T_c , are small because they involve pairing of charged particles, that is, electrons. However, in one dimension, because of the fact that the low-energy excitations are independent spin and charge collective modes¹⁰, superconducting pairing involves only the (neutral) spin degrees of freedom, and hence the gap can be large⁹. This would be a good starting point for a mechanism of high-temperature superconductivity, except for the fact⁷ that higher-dimensional interactions typically lead to CDW order rather than superconductivity because the CDW susceptibility is the more divergent as $T \rightarrow 0$; see equation (1). On the other hand, as we show below, transverse stripe fluctuations have a significant effect on the competition between superconducting and CDW order. Such fluctuations are unimportant in conventional quasi-one-dimensional solids, where there are no background spins to define the meandering rivers of charge, and the constituent molecules, upon which the electrons move, have a large mass. But the zero-point energy $\hbar\bar{\omega}$ of the transverse stripe fluctuations becomes a significant energy scale in a stripe phase of a doped antiferromagnet which arises from a purely electronic correlation effect. Indeed, we have found that, as $\hbar\bar{\omega}$ increases in magnitude relative to the Coulomb interaction V , there is a first-order phase transition from an electronic solid to an electronic smectic liquid crystal, in which the CDW fluctuations on neighbouring stripes are uncorrelated at large distances. At the same time, the effective Josephson coupling J between superconducting fluctuations on neighbouring stripes is greatly enhanced as a function of increasing $\hbar\bar{\omega}/V$.

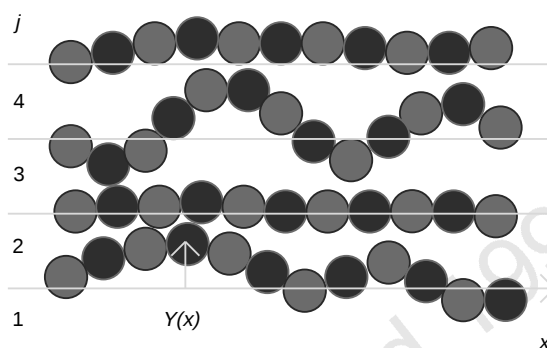


Figure 1 Schematic representation of a smectic stripe phase. The shaded circles represent periodic structures along the stripes, which are forced out of phase by the transverse fluctuations.

We begin with a simple model of a two-dimensional array of stripes along the x direction. In the ordered state, we can safely ignore dislocations and overhangs. This allows us to introduce a coordinate system in which points on the stripes are labelled by a stripe number, j , and by a position x along the stripe direction. Then the stripe configuration is described by the transverse displacement in the y direction, $Y_j(x)$, of the j th stripe at position x (Fig. 1). Because of the spin gap, the only other low-energy degrees of freedom involve fluctuations of the charge density, $\rho_j(x)$, on each stripe. However, because the stripes do not remain straight, the electronic degrees of freedom (holes) sense a fluctuating geometry. Consequently, the local charge density can be expressed as:

$$\rho_j(x) = \bar{\rho} + \rho_0 \cos[\sqrt{2\pi}\phi_j + 2k_F L_j(x)] \quad (2)$$

$$L_j(x) = \int_0^x dx' \sqrt{1 + (\partial_x Y_j)^2} + L_j(0) \quad (3)$$

where $\phi_j(x)$ defines the phase of the CDW with wavevector equal to twice the Fermi wavevector k_F and $L_j(x)$ is the arc-length, that is, the distance measured along the stripe j . The quantum dynamics of this system is equivalent to a theory of the longitudinal (ϕ_j) and transverse (Y_j) vibrations of coupled elastic strings. This defines the hamiltonian for the smectic phase. (A substrate potential would inevitably lock the smectic phase to a wavevector commensurate with the underlying crystal lattice, and hence the transverse modes would be gapped¹⁷.)

The coupling between the CDWs on neighbouring stripes is of the form:

$$H_c = \sum_j \int dx V(\Delta_j Y) \cos[\sqrt{2\pi}(\Delta_j \phi) - 2k_F(\Delta_j L)] \quad (4)$$

plus higher harmonics. Here L_j is defined in equation (3), $\Delta_j F \equiv F_{j+1} - F_j$, and the function $V[\Delta_j Y]$ reflects the fact that CDWs on adjacent stripes are more strongly coupled where the stripes are close together than when they are far apart. When this coupling is strong, it will drive the system into a fully crystalline state. Finally, there is a term in the hamiltonian representing the Josephson tunnelling of (superconducting) pairs of electrons between stripes. The tunnelling matrix element:

$$J(\Delta_j Y) \approx J_0 \exp[-\alpha \Delta_j Y] \quad (5)$$

depends roughly exponentially on the local spacing of the stripes, where α is a constant. The fact that superconductivity is a $k = 0$ order implies that the Josephson coupling does not depend on the arc length L_j , and hence it is not affected by the geometry of the stripes.

With this background, it is possible to state our central point that, to all orders in perturbation theory in powers of V , all terms that are not invariant under the transformation $\phi_j(x, t) \rightarrow \phi_j(x, t) + \delta_j$

(where t is time) for arbitrary δ_j are non-vanishing only near the 'surface', so in the thermodynamic limit the phases ϕ_j of the CDW fluctuations on neighbouring stripes are not locked together and there is no CDW order. The physical origin of this effect is easily understood. The difference in arc lengths, $\Delta_j L = L_{j+1}(x) - L_j(x)$, is a sum of contributions with random sign, which are more or less independently distributed along the distance $|x|$. For this reason, $\Delta_j L$ (and the dephasing) grow with increasing $|x|$ roughly as in a random walk, that is $|\Delta_j L|^2 \sim D|x|$, where D is a transverse diffusion constant.

This result may be obtained formally by integrating out the stripe fluctuations (Y) perturbatively in powers of V and, subsequently, J . To first order in V , the effective interaction between the CDWs on neighbouring stripes, $V^{(1)}(x; \Delta\phi(x))$, is given by the expression:

$$V^{(1)}(x; \Delta\phi) = \langle V(\Delta Y) \cos[\sqrt{2\pi}(\Delta\phi) - 2k_F(\Delta L)] \rangle \quad (6)$$

where $\langle \rangle$ implies averaging over transverse stripe fluctuations. To lowest order in a cumulant expansion:

$$V^{(1)}(x; \Delta\phi) = \bar{V}(x) \cos[\sqrt{2\pi}(\Delta\phi)] \quad (7)$$

$$\bar{V}(x) = \langle V(\Delta Y) \rangle \exp\{-(2k_F^2 \langle [\Delta L(x)]^2 \rangle)\}$$

At any non-zero temperature, in agreement with the simple physical argument given above, it is straightforward to show that $\langle [\Delta L(x)]^2 \rangle \sim A T |x|$ for large $|x|$, where A is a complicated measure of the magnitude of the transverse stripe fluctuations. It is important to note that A is dominated by short-wavelength transverse fluctuations, and is insensitive to details of their quantum dynamics. At precisely $T = 0$, for technical reasons that are not important for present purposes, the dephasing effect is somewhat more subtle, and in fact $\langle [\Delta L(x)]^2 \rangle \sim A \hbar \omega \log|x|$, which implies that $\bar{V}$ falls as a power of $|x|$ as opposed to the exponential fall-off at non-zero T . These expressions, which can readily be extended to higher order in perturbation theory and higher order in the cumulant expansion, capture the essential general point of the physics—that the effective coupling between CDWs vanishes except in a narrow 'surface' region at the ends of the stripes, and hence can be ignored in the thermodynamic limit.

The effect of Josephson coupling between stripes may be analysed in the same way. To first order in J , the effective action is proportional to:

$$\langle J \rangle \approx J_0 \exp\{(\alpha^2/2) \langle [\Delta_j Y]^2 \rangle\} \quad (8)$$

Hence the superconducting coupling is strongly enhanced by the transverse stripe fluctuations. (There is a similar enhancement of the CDW coupling, V , but it is overwhelmed by the dephasing effect.) Physically, this enhancement reflects the fact that the mean value of J is dominated by regions where neighbouring stripes come close together so that $\langle J \rangle$ is very much larger than the median. From equation (1) it can be seen that, when $K_c > 1/2$, the pair susceptibility on an individual stripe diverges as $T \rightarrow 0$ and hence, for non-zero J , the smectic phase is always globally superconducting below a finite (Kosterlitz–Thouless) ordering temperature, $T_c \sim (\langle J \rangle \Delta_s)^{K_c/(2K_c-1)}$. Because of the broken rotational symmetry of this phase, all that can be said about the symmetry of the superconducting state is that it is singlet; its symmetry is necessarily a mixture of 's-wave' and 'd-wave'. On the other hand, for $K_c < 1/2$ and $\langle J \rangle$ sufficiently small, the system remains a (quantum critical) non-Fermi liquid all the way to $T = 0$. Although such quantum critical phases are common in one dimension¹⁰, where they are often called 'Luttinger liquids', we believe this is the first theoretically well justified example in two dimensions.

To complete the physical picture of the quantum smectic, we construct a global phase diagram, shown schematically in Fig. 2, by considering the possible zero and finite temperature phase transitions from the smectic state to states with other symmetries. This can be done, to a large extent, on the basis of general considerations

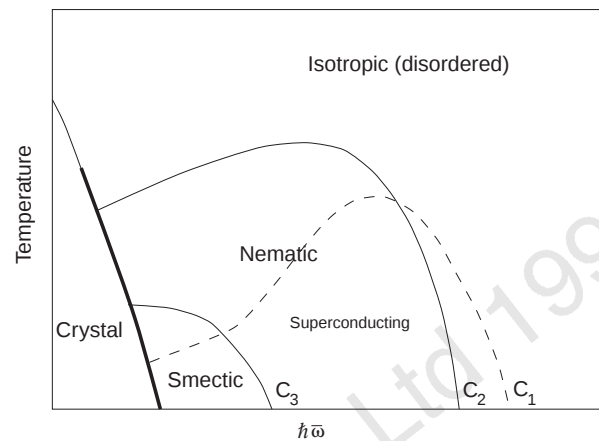


Figure 2 Phase diagram for $K_c > 1/2$. Here $\hbar\omega$ is a measure of the magnitude of the transverse zero-point fluctuations of the stripes. Thin lines represent continuous transitions, and the thick line is a first-order transition. The dashed line is the superconducting T_c . The symbols C_j label various quantum-critical points. Depending on microscopic details, the positions of C_1 and C_2 could be interchanged.

of symmetry and by analogy with the phase diagram of conventional liquid crystals; the argument relies on nothing more than the existence (and electronic character) of the quantum smectic phase.

Along the $T = 0$ axis of the figure, the phases evolve as $\hbar\omega$ is varied. Starting from the smectic phase:

(1) To the left, as the system becomes progressively more 'classical', there is a (first-order) phase transition to a crystalline state, in which the phases of the CDWs on neighbouring stripes are locked, the transverse stripe fluctuations become the phonons of a fully ordered crystal, superconducting order is destroyed, and the system becomes globally insulating.

(2) To the right, as the system becomes more quantum and, in particular, when the r.m.s. magnitude of the transverse fluctuations of the stripes becomes comparable to their spacing, there is a $T = 0$ transition to a quantum nematic phase. This transition is driven by dislocations which destroy the stripe positional ordering at long distances. We generally expect this transition to be continuous. This implies that, for $K_c > 1/2$, the superconducting order must continue across the smectic to nematic phase boundary. Similarly, in the case $K_c < 1/2$, the Luttinger liquid behaviour must persist across the phase boundary.

(3) At still larger $\hbar\omega/V$, there must be a transition to an isotropic phase. Landau theory suggests that the nematic to isotropic transition should be continuous in two spatial dimensions, although it is first-order in three.

(4) For $K_c > 1/2$, there are two possible schemes for the termination of the high-temperature superconducting order with increasing $\hbar\omega$: If the nematic region of the phase diagram is narrow, so that significant local stripe correlations survive into the isotropic phase, then the superconducting state could survive until some larger value of $\hbar\omega$, as shown in Fig. 2; in this case, the superconducting state will have a pure symmetry ('s' or 'd') where it extends into the isotropic phase. Otherwise, the high-temperature superconducting phase could terminate at a critical point within the nematic phase. In either case, beyond this point, the ground state is a (possibly) anisotropic Fermi liquid (similar to a conventional metal) or, if there remain sufficient residual interactions, a low-temperature superconductor. A highly schematic view of the local stripe order in these various phases is shown in Fig. 3.

Except for the choices among the possible schemes described above, and so long as the phase transitions are continuous, the topology of the phase diagram is constrained to be as shown in Fig. 2 for the case $K_c > 1/2$. At high enough temperature, there must be a

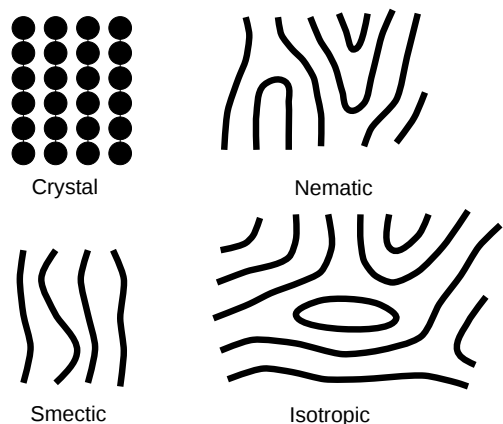


Figure 3 Schematic view of the local stripe order in the various phases discussed in the text. Here, we have assumed that the stripes maintain their integrity throughout, although in reality they must certainly become less and less well defined as the system becomes increasingly quantum, until eventually they are not the correct variables for describing the important correlations in the system. Heavy lines represent liquid-like stripes, along which the electrons can flow, whereas the filled circles represent pinned, density-wave order along the stripes. The stripes are shown executing more or less harmonic oscillations in the smectic phase. Two dislocations, which play an essential role in the smectic-to-nematic phase transition, are shown in the view of the nematic phase.

transition from a smectic to an isotropic (symmetric) phase. As at zero temperature, the smectic order is destroyed in a sequence of two transitions: (1) a dislocation unbinding transition to an “Ising nematic” phase¹⁸ which has short-range positional order but breaks four-fold rotational symmetry, and (2) a transition to the isotropic state. The superconducting T_c rises with $\hbar\omega$ through the smectic and nematic phases, reflecting the enhancement of the Josephson coupling, J , by transverse stripe fluctuations; it decreases at larger $\hbar\omega$ following the isotropic to nematic phase boundary, as we expect that the stripes lose their local integrity far into the isotropic phase. A further detail is that both the crystalline and smectic regions are actually a series of commensurate phases at $T = 0$, and a complicated pattern of commensurate and incommensurate phases for $T \neq 0$. Whereas the commensurate smectic has true positional long-range order, the incommensurate smectic will have only power-law correlations, and there is no true broken translational symmetry—only quasi-long-ranged positional order. If any of the phase transitions were discontinuous (first-order), the character of the phase diagram would change. An important and interesting possibility is that C_1 could be replaced by a line of first-order phase transitions, extending to finite temperature in either the nematic or isotropic regions of the phase diagram.

As mentioned above, crystals do not have the full rotational symmetry of free space. A crystal field with two-fold symmetry would change the nematic-to-isotropic phase transition into a crossover, which nevertheless would be quite sharp if the field is small. We note that our analysis could also be applied to systems with low-energy spin degrees of freedom by considering the most general model of the one-dimensional electron gas with or without a charge gap¹⁰.

What are the experimental signatures of the electronic liquid-crystal phases? The most direct would come from peaks in the static and dynamic, spin and charge structure factors, measured by neutron and X-ray scattering. Long-range order transverse to the stripes is indicated by a Bragg peak for which the component, q_x , of the wavevector along the stripe direction is equal to zero. There are additional peaks with $q_x \neq 0$ corresponding to CDW ordering along the stripes—Bragg peaks in the case of the crystalline phase, and power-law singularities for the smectic. In practice, the latter may be

of low intensity and difficult to observe. However, the electrical conductivity allows an unambiguous distinction to be made between the insulating crystalline phase and the metallic smectic phase. In the nematic phase near to the smectic phase boundary, sharp peaks corresponding to smectic order with a long but finite correlation length should also be observable in the static structure factor. In addition, the electronic properties should be strongly anisotropic, as this phase breaks four-fold rotational symmetry, even in a nominally tetragonal material. This analysis is complicated by the effects of quenched disorder, which always leads to a rounding of the Bragg peaks in two dimensions, even in the crystalline phase.

There is strong direct experimental evidence of electronic liquid-crystal phases in the copper oxide superconductors. Neutron-scattering experiments by Tranquada *et al.*^{11,12} have found static peaks, corresponding in incommensurate spin and charge stripe order, in $\text{La}_{1.6-x}\text{Nd}_{0.4}\text{Sr}_x\text{CuO}_4$. The stripes are along the CuO direction and the material is simultaneously a bulk superconductor. The peaks have a small but finite width which is consistent with a nematic stripe phase in an orientating potential. However, because of the presence of quenched disorder in these materials, the peaks could possibly arise from a disrupted smectic phase. In this material, the orientation of the oxygen octahedra produces a two-fold symmetry-breaking potential that drives the material either into or close to the smectic phase, and freezes the dynamics. In $\text{La}_{2-x}\text{Sr}_x\text{CuO}_4$ there are similar incommensurate peaks in the magnetic neutron-scattering factor at about the same position in $\mathbf{k}$ -space, but they are inelastic^{13–16}; that is, there are dynamically fluctuating analogues of the stripe phases seen in $\text{La}_{1.6-x}\text{Nd}_{0.4}\text{Sr}_x\text{CuO}_4$. Here the two-fold lattice potential is, itself, dynamical. Neutron-scattering experiments on underdoped $\text{YBa}_2\text{Cu}_3\text{O}_{7-\delta}$ also have found dynamical incommensurate peaks^{19,20}, corresponding to low-energy dynamical stripe fluctuations. □

Received 21 July 1997; accepted 31 March 1998.

- Kivelson, S. A. & Emery, V. J. Topological doping. *Synth. Met.* **80**, 151–158 (1996).
- Zaanen, J. & Gunnarsson, O. Charged magnetic domain lines and the magnetism of high T_c oxides. *Phys. Rev. B* **40**, 7391–7394 (1989).
- Schulz, H. J. Incommensurate antiferromagnetism in the 2-dimensional Hubbard model. *Phys. Rev. Lett.* **64**, 1445–1448 (1990).
- Kivelson, S. A. & Emery, V. J. in *Strongly Correlated Electron Materials: The Los Alamos Symposium 1993* (eds Bedell, K. S., Wang, Z., Meltzer, D. E., Balatsky, A. V. & Abrahams, E.) 619–650 (Addison-Wesley, Redwood City, 1994).
- Nayak, C. & Wilczek, F. Populated domain walls. *Phys. Rev. Lett.* **78**, 2465–2468 (1997).
- Castro-Neto, A. H. Superconducting phase coherence in striped cuprates. *Phys. Rev. Lett.* **78**, 3931–3934 (1997).
- Devreese, J. T., Evrard, R. P. & van Doren, V. E. (eds) *Highly Conducting One-Dimensional Solids* (Plenum, New York, 1979).
- Chaikin, P. M. & Lubensky, T. C. *Principles of Condensed Matter Physics* (Cambridge Univ. Press, 1995).
- Emery, V. J., Kivelson, S. A. & Zachar, O. Spin-gap proximity effect mechanism of high-temperature superconductivity. *Phys. Rev. B* **56**, 6120–6147 (1997).
- Emery, V. J. in *Highly Conducting One-Dimensional Solids* (eds Devreese, J. T., Evrard, R. P. & van Doren, V. E.) 247–303 (Plenum, New York, 1979).
- Tranquada, J. M. *et al.* Evidence for stripe correlations of spins and holes in copper oxide superconductors. *Nature* **375**, 561–563 (1995).
- Tranquada, J. M. *et al.* Coexistence of, and competition between, superconductivity and charge-stripe order in $\text{La}_{1-x}\text{Nd}_x\text{Sr}_x\text{CuO}_4$. *Phys. Rev. Lett.* **78**, 338–341 (1997).
- Cheong, S.-W. *et al.* Incommensurate magnetic fluctuations in $\text{La}_{1-x}\text{Sr}_x\text{CuO}_4$. *Phys. Rev. Lett.* **67**, 1791–1794 (1991).
- Mason, T. E., Aeppli, G. & Mook, H. A. Magnetic dynamics of superconducting $\text{La}_{1.86}\text{Sr}_{0.14}\text{CuO}_4$. *Phys. Rev. Lett.* **68**, 1414–1417 (1992).
- Thurston, T. R. *et al.* Low-energy incommensurate spin excitations in superconducting $\text{La}_{1.85}\text{Sr}_{0.15}\text{CuO}_4$. *Phys. Rev. B* **46**, 9128–9131 (1992).
- Yamada, K. *et al.* Direct observation of a magnetic gap in superconducting $\text{La}_{1.85}\text{Sr}_{0.15}\text{CuO}_4$ ($T_c = 37.3\text{ K}$). *Phys. Rev. Lett.* **75**, 1626–1629 (1995).
- Pokrovsky, V. L., Talapov, A. L. & Bak, P. in *Solitons* (eds Trullinger, S. E., Zakharov, V. E. & Pokrovsky, V. L.) 71–127 (North Holland, Amsterdam, 1986).
- Abanov, A. *et al.* Phase diagram of ultrathin ferromagnetic films with perpendicular anisotropy. *Phys. Rev. B* **51**, 1023–1038 (1995).
- Tranquada, J. M. Charge stripes and spin correlations in copper oxide superconductors. *Physica C* **282–287**, 166–169 (1997).
- Dai, P., Mook, H. A. & Dogan, F. Incommensurate magnetic fluctuations in $\text{YBa}_2\text{Cu}_3\text{O}_{6.6}$. *Phys. Rev. Lett.* **80**, 1738–1741 (1998).

Acknowledgements. We thank J. Tranquada for discussions and V. Pokrovskii for comments. This work was supported in part by the NSF at UCLA, UIUC and ITP-UCSB; and in part by the Division of Materials Science, US DOE, at Brookhaven. Two of us (E.E. and S.K.) were participants at the ITP Program on Quantum Field Theory in Low Dimensions.

Correspondence and requests for materials should be addressed to V.J.E. (e-mail: emery@cmth.phy.bnl.gov).

Signatures of distinct dynamical regimes in the energy landscape of a glass-forming liquid

Srikanth Sastry^{*†}, Pablo G. Debenedetti^{*} & Frank H. Stillinger^{‡§}

^{*} Department of Chemical Engineering, [‡] Princeton Materials Institute, Princeton University, Princeton, New Jersey 08544, USA

[§] Bell Laboratories, Lucent Technologies, Murray Hill, New Jersey 07974, USA

[†] Permanent address: Jawaharlal Nehru Center for Advanced Scientific Research, Bangalore 560064, India.

Most materials attain a glassy state at low temperatures under suitable methods of preparation. This state exhibits the mechanical properties of a solid, but shows microscopic structural disorder^{1,2}. A comprehensive understanding of the glassy state is, however, still lacking³. A widespread assumption is that the non-

exponential relaxation processes observed in the dynamics of glasses—and also in protein dynamics, protein folding and population dynamics—are (in common with other manifestations of complex dynamics) strongly influenced by the underlying energy landscape associated with the structural configurations that the system may adopt. But concrete evidence for this in studies of glass formation has been scarce. Here we present such evidence, obtained from computer simulations of a model glass-forming liquid. We demonstrate that the onset of non-exponential relaxation corresponds to a well defined temperature below which the depth of the potential-energy minima explored by the liquid increases with decreasing temperature, and above which it does not. At lower temperatures, we observe a sharp transition when the liquid gets trapped in the deepest accessible energy basin. This transition temperature depends on the cooling rate, in a manner analogous to the experimental glass transition. We also present evidence that the barrier heights separating potential-energy minima sampled by the liquid increase abruptly at a temperature above the glass transition but well below the onset of non-exponential relaxation. This identification of a relationship between static, topographic features of the energy landscape and complex dynamics holds

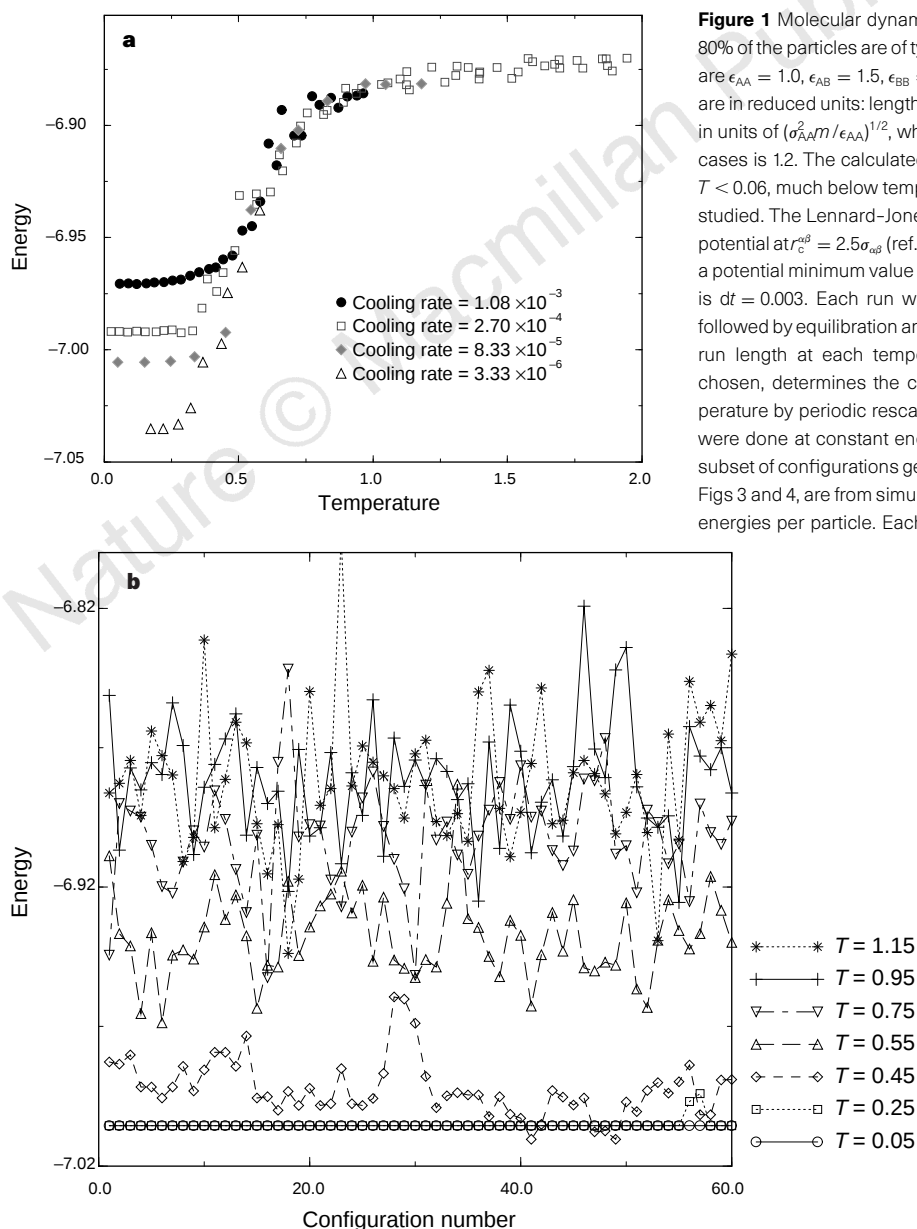


Figure 1 Molecular dynamics simulations of a binary Lennard-Jones mixture²¹. 80% of the particles are of type A, 20% are type B, and Lennard-Jones parameters are $\epsilon_{AA} = 1.0$, $\epsilon_{AB} = 1.5$, $\epsilon_{BB} = 0.5$, $\sigma_{AA} = 1.0$, $\sigma_{AB} = 0.8$ and $\sigma_{BB} = 0.88$. All quantities are in reduced units: length in units of σ_{AA} , temperature in units of ϵ_{AA}/k_B , and time in units of $(\sigma_{AA}^2 m / \epsilon_{AA})^{1/2}$, where m is the mass of the particles. The density ρ in all cases is 1.2. The calculated pressures for the system remain positive except for $T < 0.06$, much below temperatures where the system forms a glass in all cases studied. The Lennard-Jones potential, with a quadratic cut-off and shifting of the potential at $r_c^{\alpha\beta} = 2.5\sigma_{\alpha\beta}$ (ref. 29), $\alpha, \beta \in A, B$ is used. Our cut-off procedure results in a potential minimum value which is $\sim 4\%$ smaller than that in ref. 21. The time step is $\Delta t = 0.003$. Each run was initialized by equilibration at a high temperature, followed by equilibration and data collection runs at a series of temperatures. The run length at each temperature, together with the number of temperatures chosen, determines the cooling rate. Equilibration was done at constant temperature by periodic rescaling of the velocities of particles. Data collection runs were done at constant energy. Local-energy minimization was performed for a subset of configurations generated at each temperature. Data, as well as those in Figs 3 and 4, are from simulations of 256 particles. **a**, The average of the minimum energies per particle. Each data point represents an average over 100 inherent structures, except for cooling rate 8.33×10^{-5} , where 60 configurations were used. The three lower cooling-rate data were obtained from a single run; for the highest cooling rate, the data were averaged over four different cooling runs (25 inherent structures were sampled in each case) to reduce the noise. For temperatures $T = 3.0$, 4.0 and 5.0, the average energies are -6.868 , -6.872 and -6.867 respectively, confirming that at high temperatures ($T > 1.0$) the inherent-structure energies reach a constant value independent of temperature. **b**, Individual minimum energies for the configurations at cooling rate 8.33×10^{-5} . At high and intermediate temperatures, these individual energies cover a broad range, and only show a gradual trend towards lower values as the temperature is lowered. A qualitative difference is apparent at low temperatures, with the sampled energies becoming narrowly distributed around the average values. The statistical independence of individual points, however, varies strongly with temperature, particularly as these data represent a single cooling run.

the promise of a clearer, possibly thermodynamic, understanding of the glass transition.

Glasses can be formed in numerous ways⁴. The term 'glass transition', however, is commonly used to designate the disappearance of structural relaxation in a liquid on cooling to low temperatures; the material then becomes rigid, while retaining microscopic structural disorder. Although it is conventional to speak of a glass transition temperature, T_g , the experimentally determined T_g of a material depends on how fast the liquid is cooled¹. The possible existence of an underlying thermodynamic transition remains open, and a topic of great interest^{5–7}. At low temperatures, in addition to a rapid decrease of relaxation rates, a glass-forming liquid exhibits 'complex dynamics' such as non-exponential relaxation², breakdown of the Stokes–Einstein relation^{8,9} and translation–rotation decoupling¹⁰. The microscopic disorder, complex dynamics and loss of relaxation at finite temperature observed in glass-forming liquids find analogies in various other 'complex systems', most notably spin glasses¹¹.

An appealing approach for understanding complex dynamics is to consider the influence of a system's 'energy landscape' on the relaxation processes it displays^{12–14}. The dynamics of the system is viewed as the motion of the 'state point' (described by the coordinates of all particles) in the $3N$ -dimensional configuration space, where N is the number of particles. The potential energy of the system, a function of particle coordinates, defines a complicated $3N$ -dimensional surface or 'landscape'. We may partition the configuration space into 'basins', such that a local minimization of the potential energy maps any point in a basin to the same minimum. The properties of the system at a given temperature are dictated by the basins sampled and their mutual accessibility. At high temperatures, kinetic energy permits access to most basins. At lower temperatures, the sampling shifts to lower energies and mutual access among basins becomes subject to considerable 'activation'.

A kinetic description of glassy dynamics that has been much studied, mode-coupling theory (MCT)^{15–17}, makes no formal con-

tact with the energy landscape picture. In its idealized version, MCT predicts a critical temperature, T_c , where dynamical quantities diverge. It has subsequently been shown¹ that T_c lies above T_g , its singular character arising from approximations of the idealized theory. It has been suggested¹⁸ that activated dynamics, unaccounted for in MCT, begins to be dominant near T_c (we use the term 'activated relaxation' here as commonly used in the supercooled liquids literature, for example in ref. 18, to denote relaxation dominated by episodic particle motions that require overcoming appreciable energy or activation barriers). Further, it has been suggested that T_c demarcates temperatures where the system explores deeper regions of the potential-energy surface from those at which it has access to all regions¹⁹. But despite the importance of the landscape paradigm, quantitative measures of the manner in which a liquid samples the potential-energy landscape are scarce.

We study a binary mixture of particles interacting via the Lennard–Jones potential, originally proposed as a model for $\text{Ni}_{80}\text{P}_{20}$ (ref. 20), and widely used as a model glass-former in computer simulations^{21,22}. We perform molecular dynamics simulations at constant volume, from very high to deeply supercooled temperatures, each set of simulations being carried out at a given cooling rate. We perform a local potential-energy minimization for selected configurations. The resulting energy-minimum configurations (called inherent structures²³) serve as markers of the configuration space explored by the system at any given temperature. Although most experimental studies of glass formation are done at constant pressure, the phenomenology has been observed to remain unchanged when volume is held constant instead^{24,25}.

In Fig. 1a we show the temperature-dependent average energies of the inherent structures, for four different cooling rates. At high temperature ($T > 1.0$), the energies do not change significantly with temperature. Between $T \approx 1.0$ and $T = 0.3–0.4$, the energies decrease progressively. Below this range of temperatures, the energies are constant once again. Figure 1b shows that at high temperature the system explores a broad range of minimum energies and

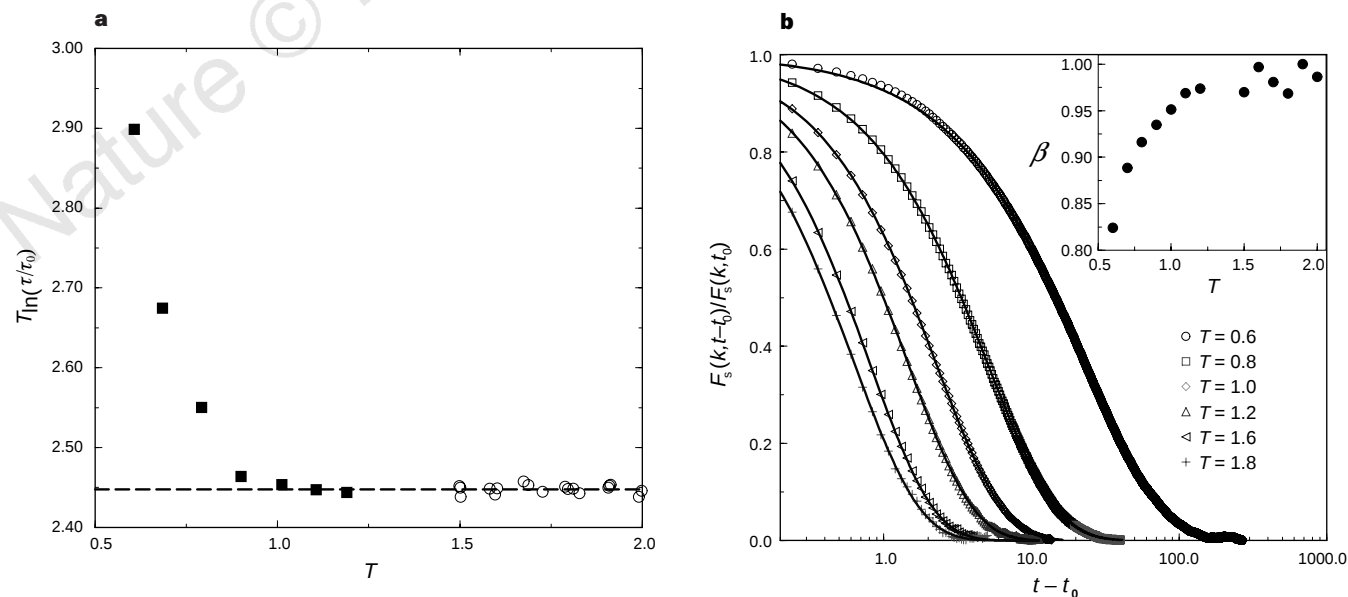


Figure 2 Transition from high-temperature behaviour to the 'landscape-influenced' regime where non-exponential relaxation sets in. **a**, Relaxation times from the space Fourier transform $F_s(k, t)$ of the self part of the van Hove correlation function (self intermediate scattering function). The wavevector is $k = 7.21\sigma_{AA}^{-1}$, close to the first peak of the static structure factor. The infinite temperature relaxation time τ_0 is obtained by fitting $\tau(T)$ values for $T = 1.5–2.0$ (circles) to an Arrhenius form. The quantity plotted, $T \ln(\tau/\tau_0)$, is constant when τ displays Arrhenius behaviour. Deviation from the constant, high-temperature value is seen around $T = 1.0$

(squares). **b**, The self intermediate scattering function displayed with a shift in the time origin to $t_0 = 1.2$, and normalized to the value at t_0 . We use this transformation as a convenient procedure to eliminate the gaussian time dependence at short t (ref. 26). The fits are to the stretched exponential form in equation (2). The inset shows the exponent values $\beta(T)$. As a check, we fitted our data to a simple exponential form, and confirmed that above $T = 1.0$ fits were satisfactory, whereas below $T = 1.0$, the fits showed systematic deviations indicating slower than exponential decay. Data are from simulations of 1,372 particles.

that on lowering the temperature, the sampling becomes progressively biased towards lower energies, becoming more narrowly distributed around the average values.

We expect a qualitative change in the dynamics below the crossover at around $T = 1.0$, based on the observed T -dependence of the inherent-structure energies. To evaluate this expectation, we calculate the self intermediate scattering function $F_s(k, t)$ (ref. 26), the Fourier transform at wavevector k of the van Hove self-correlation function

$$G_s(r, t) = \frac{1}{N} \sum_{i=1}^N \langle \delta(|\mathbf{r}_i(t) - \mathbf{r}(0)| - r) \rangle \quad (1)$$

which describes the probability of finding at time t a particle at distance r from its location at $t = 0$. We define a relaxation time $\tau(T)$ as the time when $F_s(k, t) = 1/e$. At high temperature, $\tau(T)$ displays the Arrhenius form, $\tau = \tau_0 \exp(E/k_B T)$ (k_B is the Boltzmann constant and E is the activation energy, a constant). As shown in Fig. 2a, this simple behaviour breaks down around $T = 1.0$, below which the measured τ are progressively higher than extrapolations from higher temperatures. This behaviour corresponds to that of 'fragile liquids' in the strong–fragile classification¹⁸. At high temperatures, the time dependence of $F_s(k, t)$ is exponential, whereas at lower temperatures, the Kohlrausch–Williams–Watts stretched exponential form

$$F_s(k, t) = \exp \left[- \left(\frac{t}{\tau(T)} \right)^{\beta(T)} \right] \quad (2)$$

provides a good fit to the long-time behaviour. In equation (2) $\beta(0 < \beta < 1)$ is a number that quantifies the deviation from strict exponential behaviour. We consider the functions $[F_s(k, t - t_0)] / (F_s(k, t_0), t > t_0$, which we fit to the form in equation (2) to quantify their long-time behaviour. The fits and the values for $\beta(T)$ are shown in Fig. 2b, which shows that $\beta(T)$ drops considerably below 1 as the temperature decreases below $T = 1.0$. We call this the 'landscape-influenced' regime. These results demonstrate the connection between the onset of 'slow dynamics' and a change in the manner of exploration of the potential-energy landscape. Such a connection has been proposed before^{4,12–14}, but our results, we believe, constitute its first clear demonstration.

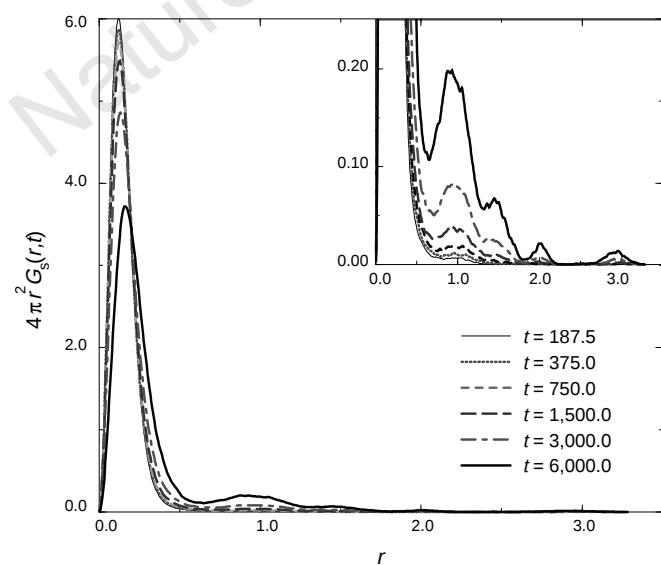


Figure 3 The van Hove self-correlation function for $T = 0.425$, shown for different values of t . Inset shows a magnification of the secondary peaks corresponding to activated events. Such secondary peaks have previously not been observed for $T > 0.466$ in the model we study²¹, but have been observed in a related model²⁸ at low temperatures.

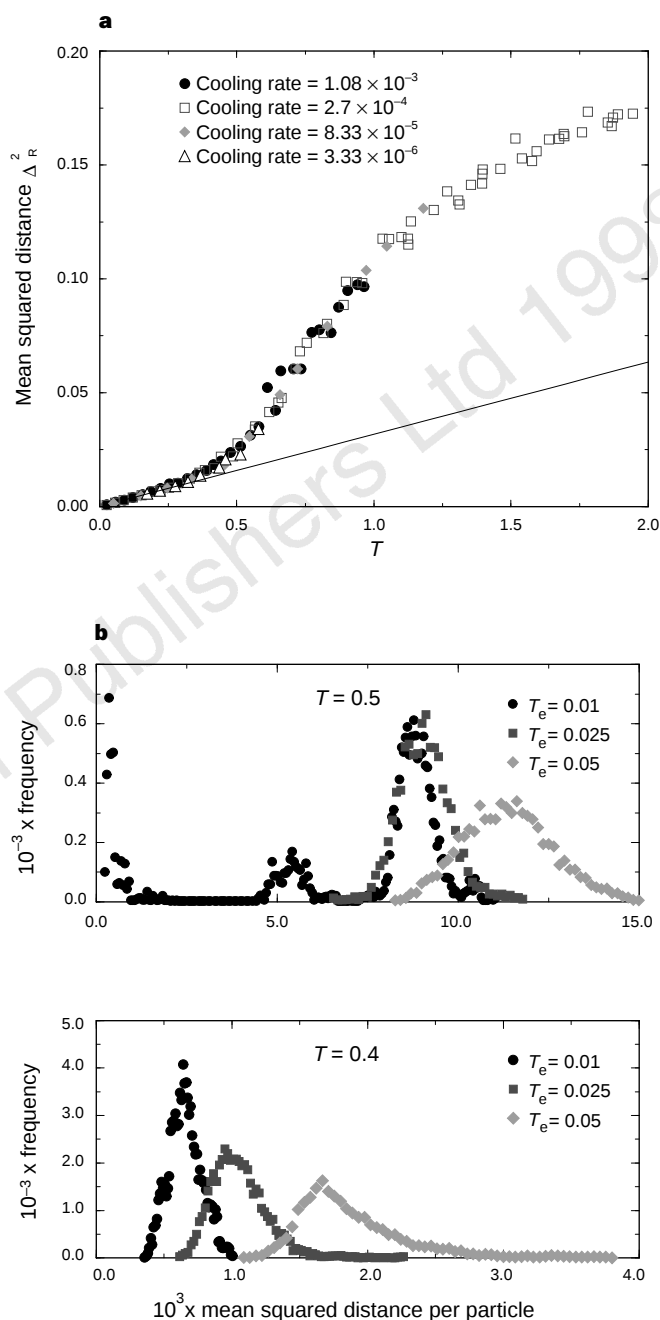


Figure 4 Onset of the 'landscape-dominated' regime at $T \approx 0.45$. **a**, The average mean-squared distance per particle Δ_s^2 between a typical liquid configuration and the corresponding inherent structure. For a quadratic minimum, Δ_s^2 is equivalent to average squared displacements due to harmonic vibrations, and hence proportional to the temperature. The reference straight line, whose slope equals that of mean-squared distances for harmonic vibrations around an inherent structure for $T = 0.175$, cooling rate $= 3.33 \times 10^{-6}$, suggests that at low temperatures Δ_s^2 indeed arises from harmonic vibrations. The observed sharp deviation from the low-temperature straight line thus suggests the onset of anharmonicity in the sampled basins. **b**, Distribution of squared distances sampled when an inherent structure is excited to energies labelled by T_e , the temperature corresponding to the kinetic energy of excitation. The subscript 'e' distinguishes this 'excitation' temperature from that of the liquid from which the inherent structures are obtained. We note that all values of T_e are substantially lower than the glass transition temperature T_g . For inherent structures obtained from the higher temperature $T = 0.5$, the system finds basins of nearby energy minima for all but the lowest level of excitation. For the lower temperature $T = 0.4$, this is not the case, and the system remains confined to the basin of the initial inherent structure.

Examination of the internal-energy and pressure curves for the liquid from which the inherent structures are obtained reveals that the sharp low-temperature break in the inherent-structure energy curve corresponds to T_g as commonly estimated in simulations²⁷. Thus, calculation of inherent-structure energies offers an alternative way for locating the glass transition temperature in computer simulations. The advantage of our procedure is that the signature of the transition is very sharp, although it requires considerable additional computation. Based on the differences between curves for successive cooling rates, we note that the system begins to fall out of equilibrium well above the low-temperature plateau.

The temperature range slightly above T_g , where inherent-structure energies show a gradual change with temperature, includes the temperature where MCT predicts a divergence in relaxation times ($T_c \approx 0.435$ in ref. 21). Instead of this divergence, one observes deviations from MCT predictions as T_c is approached, possibly due to 'activated processes'^{12,13,18}. These lead to secondary peaks in the van Hove self-correlation function^{21,28}, resulting from rare jumps of particles over distances roughly equal to interparticle separations. Figure 3 shows $G_s(r, t)$ for $T = 0.425$. Secondary peaks appear at positions $r \approx 1.0, 1.44, 2.0$ and 3.0 , whereas no such distinct peaks are found at higher temperatures. Thus, at $T = 0.425$, the system shows evidence of activated dynamics. We call this the 'landscape-dominated' regime.

Figure 4a shows the mean-squared distance Δ_R^2 between a typical liquid configuration and the corresponding energy minimum, defined as $\Delta_R^2 = (1/N)\sum(\mathbf{r}_i - \mathbf{r}_i^{(q)})^2$, where $\mathbf{r}_i$ are the coordinates of particles i , and $\mathbf{r}_i^{(q)}$ are coordinates in the corresponding inherent structures. A crossover from roughly linear behaviour at low temperatures to a more rapid increase occurs at around $T = 0.45$. In addition, the dependence on Δ_R^2 of the path length traversed during energy minimization also displays a break at $T \approx 0.45$ (data not shown), indicating an increased roughness of the landscape sampled at higher temperatures. These observations indicate a sharp change in the local topography of the sampled landscape around $T = 0.45$, which is close to the previous estimate of T_c (ref. 21).

Further insight is obtained by exciting inherent structures obtained from a given temperature to various energies (with kinetic energies corresponding to very low excitation temperatures, T_e) and monitoring the distribution of distances from the starting inherent-structure configuration, which are shown in Fig. 4b. At $T = 0.4$, the location of the peaks of the distribution is roughly proportional to T_e . At $T = 0.5$, on the other hand, even at very low excitation, the system finds other basins, as evidenced by multiple peaks and substantially higher sampled distances. This indicates that whereas at high temperature the system explores a part of the landscape with low barriers between energy minima, at lower temperatures it explores minima with substantially higher energy barriers. □

Received 29 December 1997; accepted 1 April 1998.

1. Debenedetti, P. G. *Metastable Liquids* (Princeton Univ. Press, Princeton, 1996).
2. Ediger, M. D., Angell, C. A. & Nagel, S. R. Supercooled liquids and glasses. *J. Phys. Chem.* **100**, 13200–13212 (1996).
3. Anderson, P. W. Through the glass lightly. *Science* **267**, 1615 (1995).
4. Angell, C. A. Formation of glasses from liquids and biopolymers. *Science* **267**, 1924–1935 (1995).
5. Kauzmann, W. The nature of the glassy state and the behavior of liquids at low temperatures. *Chem. Rev.* **43**, 219–256 (1948).
6. Gibbs, J. H. & DiMarzio, E. A. Nature of the glass transition and the glassy state. *J. Chem. Phys.* **28**, 373–383 (1958).
7. Stillinger, F. H. Supercooled liquids, glass transitions, and the Kauzmann paradox. *J. Chem. Phys.* **88**, 7818–7825 (1988).
8. Heuberger, G. & Silescu, H. Size dependence of tracer diffusion in supercooled liquids. *J. Phys. Chem.* **100**, 15255–15260 (1996).
9. Chang, I., Fujara, F., Geil, B., Heuberger, G. & Silescu, H. Translational and rotational molecular motion in supercooled liquids studied by NMR and forced Rayleigh scattering. *J. Non-Cryst. Solids* **172–174**, 248–255 (1994).
10. Cicerone, M. T., Blackburn, F. R. & Ediger, M. D. How do molecules move near T_g ? Molecular rotation of six probes in *o*-terphenyl across 14 decades in time. *J. Chem. Phys.* **102**, 471–479 (1995).
11. Parisi, G. New ideas in glass transitions. in *Proc. Int. Workshop on The Morphology and Kinetics of Phase Separating Complex Fluids* (eds Chen, S. H., Mallamace, F. & Tartaglia, P.) *Il Nuovo Cimento D* (in the press).
12. Goldstein, M. Viscous liquids and the glass transition: a potential energy barrier picture. *J. Chem. Phys.* **51**, 3728–3739 (1969).

13. Goldstein, M. Viscous liquids and the glass transition. VII. Molecular mechanisms for a thermodynamic second order transition. *J. Chem. Phys.* **67**, 2246–2253 (1977).
14. Stillinger, F. H. A topographic view of supercooled liquids and glass formation. *Science* **267**, 1935–1939 (1995).
15. Götze, W. & Sjögren, L. Relaxation processes in supercooled liquids. *Rep. Prog. Phys.* **55**, 241–376 (1992).
16. Bengtzelius, U., Götze, W. & Sjölander, A. Dynamics of supercooled liquids and the glass transition. *J. Phys. C* **17**, 5915–5934 (1984).
17. Leutheusser, E. Dynamical model for the liquid-glass transition. *Phys. Rev. A* **29**, 2765–2773 (1984).
18. Angell, C. A. Perspective on the glass transition. *J. Phys. Chem. Sol.* **49**, 863–870 (1988).
19. Angell, C. A. in *Complex Behavior in Glassy Systems* (eds Rubi, M. & Perez-Vicente, C.) 1–21 (Springer, Berlin, 1997).
20. Weber, T. A. & Stillinger, F. H. Local order and structural translations in amorphous metal-metalloid alloys. *Phys. Rev. B* **31**, 1954–1963 (1985).
21. Kob, W. & Andersen, H. C. Testing mode-coupling theory for a supercooled binary Lennard-Jones mixture: the van Hove correlation function. *Phys. Rev. E* **51**, 4626–4641 (1995).
22. Donati, C., Douglas, J. F., Plimpton, S. J., Poole, P. H. & Glotzer, S. C. String-like cooperative motion in a supercooled liquid. *Phys. Rev. Lett.* **80**, 2338–2342 (1998).
23. Stillinger, F. H. & Weber, T. A. Packing structures and transitions in liquids and solids. *Science* **225**, 983–989 (1984).
24. Angell, C. A. *et al.* in *Slow Dynamics in Condensed Matter: Proc. 1st Tohwa Univ. Int. Symp.* (eds Kawasaki, K., Kawakatsu, T. & Tokuyama, M.) 3–19 (AIP Conf. Proc. 316, Am. Inst. Physics, New York, 1992).
25. Colucci, D. M. *et al.* Isochoric and isobaric glass formation: similarities and differences. *J. Polym. Sci. B: Polym. Phys.* **35**, 1561–1573 (1997).
26. Hansen, J.-P. & McDonald, I. R. *Theory of Simple Liquids* (Academic, London, 1986).
27. Fox, J. R. & Andersen, H. C. Molecular dynamics simulations of a supercooled monoatomic liquid and glass. *J. Phys. Chem.* **88**, 4019–4027 (1984).
28. Wahnström, G. Molecular-dynamics study of a supercooled two-component Lennard-Jones system. *Phys. Rev. A* **44**, 3752–3764 (1991).
29. Stoddard, S. D. & Ford, J. Numerical experiments on the stochastic behavior of a Lennard-Jones gas system. *Phys. Rev. A* **8**, 1504–1512 (1973).

Acknowledgements. We thank C. A. Angell, S. Glotzer, F. Sciortino and H. E. Stanley for discussions. This work was supported by the US Department of Energy and the Petroleum Research Fund.

Correspondence and requests for materials should be addressed to P.G.D. (e-mail: pdebene@pucc.princeton.edu).

Millennial-scale changes in North Atlantic circulation since the last glaciation

Thomas M. Marchitto Jr*, William B. Curry† & Delia W. Oppo†

* Massachusetts Institute of Technology/Woods Hole Oceanographic Institution Joint Program in Oceanography, Woods Hole, Massachusetts 02543, USA

† Woods Hole Oceanographic Institution, Woods Hole, Massachusetts 02543, USA

Ocean circulation is closely linked to climate change on glacial-interglacial and shorter timescales. Extensive reorganizations in the circulation of deep and intermediate-depth waters in the Atlantic Ocean have been hypothesized for both the last glaciation^{1–6} and the subsequent Younger Dryas cold interval^{3,6–10}, but there has been little palaeoceanographic study of the subtropical gyres^{11–13}. These gyres are the dominant oceanic features of wind-driven circulation, and as such they reflect changes in climate and are a significant control on nutrient cycling and, possibly, atmospheric CO₂ concentrations. Here we present Cd/Ca ratios in the shells of benthic foraminifera from the Bahama banks that confirm previous suggestions^{11,12} that nutrient concentrations in the North Atlantic subtropical gyre were much lower during the Last Glacial Maximum than they are today (up to 50% lower according to our data). These contrasting nutrient burdens imply much shorter residence times for waters within the thermocline of the Last Glacial Maximum. Below the glacial thermocline, nutrient concentrations were reduced owing to the presence of Glacial North Atlantic Intermediate Water. A high-resolution Cd/Ca record from an intermediate depth indicates decreased nutrient concentrations during the Younger Dryas interval as well, mirroring opposite changes at a nearby deep site^{3,9}. Together, these observations suggest that the formation of deep and intermediate waters—North Atlantic Deep Water and Glacial North

Atlantic Intermediate Water, respectively—wax and wane alternately on both orbital and millennial timescales.

The flanks of Little and Great Bahama banks border the North-west Providence Channel ($\sim 26^\circ\text{N}$, 78°W), which connects the North Atlantic basin (Sargasso Sea) to the Florida Straits. The waters within this channel are derived from the Sargasso Sea, and can be roughly divided into two water masses: the main thermocline of the North Atlantic subtropical gyre ($<1,000\text{ m}$ depth) and the upper (Labrador Sea) component of North Atlantic Deep Water (NADW) ($\sim 1,000\text{--}2,000\text{ m}$ depth) (ref. 12). The gyre's upper thermocline waters outcrop (intersect with the base of the mixed layer) at latitudes to the north, where Ekman pumping and buoyancy flux subduct waters along deepening isopycnals to lower latitudes, following the anticyclonic circulation of the subtropical gyre. Along this route, productivity in overlying surface waters adds decaying organic matter and nutrients. Waters near the base of the thermocline outcrop farther north (above $\sim 50^\circ\text{N}$) in a region of positive wind stress curl, Ekman upwelling, and poor ventilation. These waters circulate in the subpolar gyre before entering the subtropical gyre¹⁴ and are therefore higher in nutrients and depleted in O_2 , creating a phosphate (PO_4^{3-}) maximum and O_2 minimum near 800 m depth at Bahama banks¹¹. Deeper waters are largely derived from deep convection in the Labrador Sea¹⁵.

Dissolved cadmium has an oceanic distribution that is very similar to that of phosphorus (ref. 16). Typically, both elements are absent in surface waters, increase rapidly to shallow maxima ($\sim 800\text{--}1,000\text{ m}$ depth), then decrease slightly in deeper waters. Because the Cd content of overlying water is recorded by the Cd/Ca ratio of benthic foraminifera shells¹⁷, the nutrient content of water masses can be estimated during the past, both in terms of Cd

concentration and inferred P concentration (via the modern Cd:P relationship)^{16,18}. This method was first put to use in showing that the strength of NADW (which is relatively low in Cd and P) was substantially decreased during severe glaciations¹. Here we use glacial and Holocene Cd/Ca data from the Bahama banks to investigate the nutrient contents and circulation histories of the North Atlantic subtropical gyre and upper deep waters. Previous $\delta^{13}\text{C}$ work implied that Bahamian Last Glacial Maximum (LGM) waters had $\sim 30\text{--}60\%$ less PO_4^{3-} than today^{11,12}. These results were disputed, however, because in addition to reflecting reduced nutrient levels, high $\delta^{13}\text{C}$ values can result from colder temperatures of air-sea CO_2 exchange¹⁹. Cd/Ca ratios have the advantage of being unaffected by such thermodynamic artefacts.

A suite of 15 sediment cores and grab samples span the depth range $334\text{--}1,468\text{ m}$ on the Bahama banks. Previous $\delta^{18}\text{O}$ measurements on individual benthic foraminifera shells identified the late Holocene and LGM sections of each core¹². We picked samples of the aragonitic benthic foraminifer *Hoeglundina elegans* from these two time periods. This species faithfully records bottom-water Cd concentrations with an apparent partition coefficient $D_p = [(\text{Cd}/\text{Ca})_{\text{foram}}/(\text{Cd}/\text{Ca})_{\text{water}}] \approx 1.0$ (ref. 18). This D_p shows little or no dependence on water depth, in contrast to calcitic species which vary between 1.3 and 2.9 (ref. 20). Shells of *H. elegans* are also less prone to contamination than calcitic species, mainly because they appear to be immune to MnCO_3 overgrowths which add sedimentary Cd that is not effectively removed by cleaning¹⁸.

A total of 54 late Holocene Cd/Ca measurements were made on *H. elegans* from the 15 core-tops. The resulting water-column profile of the inferred seawater Cd concentration, Cd_w , follows the predicted profile (based on seawater $[\text{PO}_4^{3-}]$ measurements) quite closely (Fig. 1). Mean Cd_w values range from 0.04 nmol kg^{-1} at 334 m depth to a maximum of 0.42 nmol kg^{-1} at 865 m , with concentrations $\sim 0.35\text{ nmol kg}^{-1}$ below $1,000\text{ m}$ depth. A total of 22 Cd/Ca measurements were made on *H. elegans* from the LGM sections of six cores, from 303 to $1,200\text{ m}$ palaeodepth (120 m was subtracted from the present core depths to account for the glacial sea level drop)²¹. Below 600 m , LGM Cd_w averaged 0.18 nmol kg^{-1} , or $\sim 0.1\text{--}0.2\text{ nmol kg}^{-1}$ lower than today, a reduction of $40\text{--}60\%$ (Fig. 1). Application of the modern global Cd:P relationship^{16,18} to our data gives PO_4^{3-} concentrations (below 600 m) of $1.5\text{--}1.6\text{ }\mu\text{mol kg}^{-1}$ during the late Holocene and $0.8\text{--}0.9\text{ }\mu\text{mol kg}^{-1}$ during the LGM, for changes of $0.6\text{--}0.8\text{ }\mu\text{mol kg}^{-1}$. These results are in good agreement with the marked LGM $\delta^{13}\text{C}$ increase and implied $0.4\text{--}0.9\text{ }\mu\text{mol kg}^{-1}$ PO_4^{3-} decrease seen previously^{11,12}. Although these PO_4^{3-} estimates are subject to some error, arising in part from regional scatter in the global Cd:P relationship²², the magnitude of the inferred LGM-Holocene shift is a reasonable approximation. The concordance between Cd and $\delta^{13}\text{C}$ suggests that there was little or no difference in the air-sea component of Bahamian $\delta^{13}\text{C}$ ($\delta^{13}\text{C}_{\text{as}}$) between the LGM and today, although air-sea exchange effects may influence the $\delta^{13}\text{C}$ record on shorter timescales.

Three main processes could have reduced nutrients in the Bahama banks glacial thermocline: (1) the disappearance of high-nutrient Antarctic Intermediate Water (AAIW), (2) decreased productivity in gyre surface waters, and (3) decreased transit times from where isopycnals outcropped. AAIW contributes $<5\%$ of Bahamian lower thermocline waters today²³, so its complete removal would account for $<10\%$ of the glacial nutrient decrease. If modern circulation rates were maintained and if glacial North Atlantic preformed PO_4^{3-} values were no lower than today, an unrealistically large ($\sim 80\%$) glacial reduction in productivity would be required to explain the low Bahamian PO_4^{3-} levels. We therefore conclude that LGM subtropical gyre waters had much shorter transit times from where isopycnals outcropped^{11,12}, primarily due to stronger winds, a southward shift of thermocline outcrop areas²⁴, and a decreased contribution from the poorly

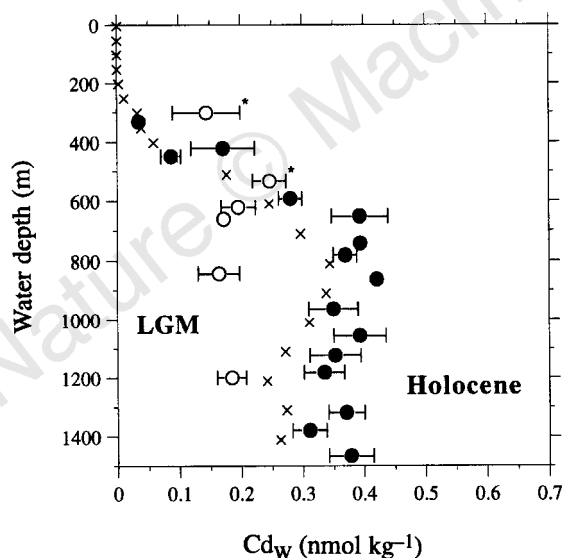


Figure 1 Vertical profiles of inferred seawater cadmium concentration, Cd_w . Values of $\text{Cd}_w = [(\text{Cd}/\text{Ca})_{\text{foram}}/D_p][\text{Ca}]_{\text{water}}$, where $[\text{Ca}]_{\text{water}} = 0.01\text{ mol kg}^{-1}$ ²⁰ are based on *H. elegans* from Late Holocene (filled circles) and LGM (open circles) sediments from the Bahama banks. Each symbol is a mean of several measurements, and error bars are standard errors ($1\sigma/n^{1/2}$). LGM depths are shifted upward by 120 m (relative to modern core depths) to account for the glacial sea level lowering²¹. Also shown are predicted modern Cd_w values (crosses) based on water-column PO_4^{3-} measurements and the global Cd:P relationship^{16,18}. Two of the 54 Holocene Cd measurements and one of the 22 LGM measurements were considered to be significantly contaminated (Cd_w much higher than expected) and were not included in the means. The two LGM data marked with asterisks are biased toward high (Holocene) concentrations because core 106GGC (534 m palaeodepth) is heavily bioturbated, and core 149JPC (303 m palaeodepth) barely penetrated glacial sediments¹².

ventilated subpolar gyre. The existence of a faster, nutrient-depleted gyre is consistent with the 'nutrient deepening' mechanism for lowering atmospheric CO_2 levels²⁵.

As waters below the modern Bahamian nutrient maximum are not strongly influenced by wind-driven circulation, their glacial nutrient reduction is best explained by changes in deep-water formation. During the LGM, formation of NADW was substantially reduced^{1,2}, being replaced by the shallower (<2,500 m) low-nutrient Glacial North Atlantic Intermediate Water (GNAIW)³⁻⁵. At 1,800 m in the Caribbean, GNAIW had a Cd_w of $0.21 \text{ nmol kg}^{-1}$ (ref. 20), about the same as the glacial Cd_w estimate at 1,200 m in the

Bahamas ($0.19 \pm 0.02 \text{ nmol kg}^{-1}$). Although it is difficult to make any clear distinction between the lower gyre and upper deep water, it is likely that the core from 965 m (OC205-2-103GGC) was at least partially influenced by the deep water. A high-resolution record of *H. elegans* Cd_w from this radiocarbon-dated core chronicles the transition from marine isotope stage 3, through the LGM, to the present (Fig. 2; see Supplementary Information for dates). Concentrations of Cd_w were relatively low during stages 3 and 2, averaging $0.17 \text{ nmol kg}^{-1}$. The deglaciation is characterized by a general increase between ~19,500 and 8,500 calendar years before present (cal. yr BP), with Holocene concentrations averaging $0.35 \text{ nmol kg}^{-1}$. Within this rise is a period of decreased Cd_w coincident with the Younger Dryas interval. Comparison to a high-resolution Cd record from the deep North Atlantic (4,450 m, Bermuda rise)^{3,9} reveals a striking inverse correlation (Fig. 3). This suggests that at times of weak NADW formation (high Cd_w at 4,450 m), GNAIW formation was strong (low Cd_w at 965 m). Another potential source of nutrient-depleted intermediate waters, Mediterranean Outflow Water, does not appear to have strengthened during either the Younger Dryas⁶ or the LGM²⁶. We

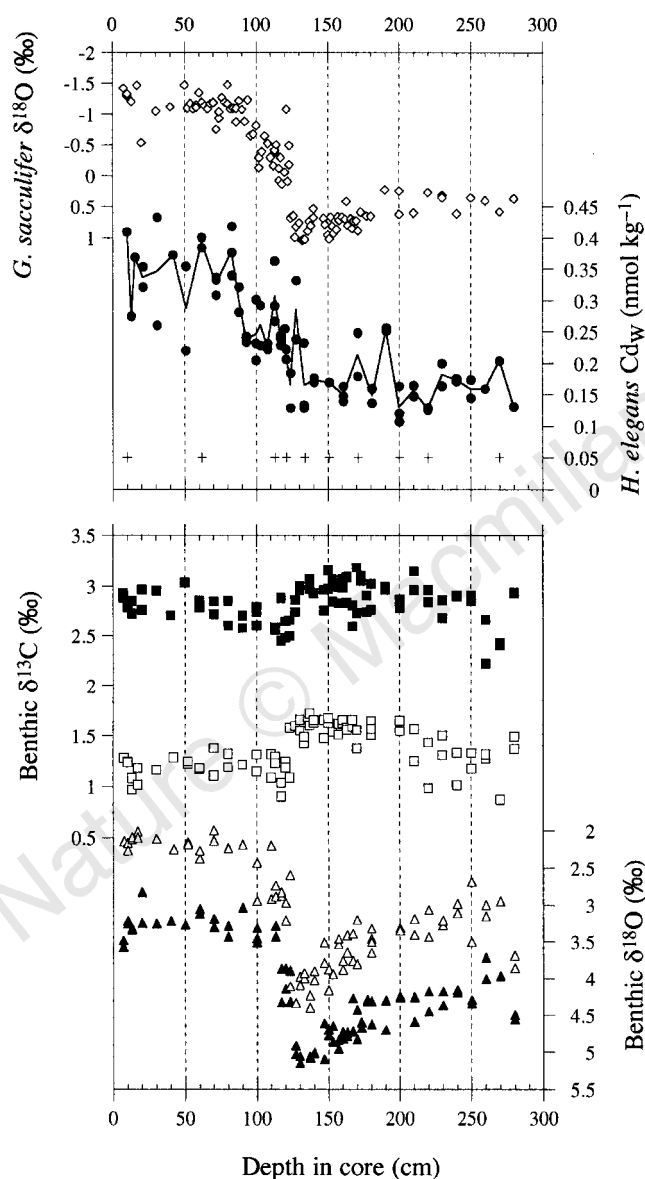


Figure 2 Data from core OC205-2-103GGC (26° 04' N, 78° 03' W; 965 m). Shown are records of planktonic $\delta^{18}\text{O}$ (*Globigerinoides sacculifer*, open diamonds), benthic Cd_w (*H. elegans*, filled circles), benthic $\delta^{13}\text{C}$ (*H. elegans*, filled squares; *Cibicides kullenbergi*, open squares), and benthic $\delta^{18}\text{O}$ (*H. elegans*, filled triangles; *C. kullenbergi*, open triangles). Crosses indicate the depths of accelerator mass spectrometer (AMS) radiocarbon dates (see Supplementary Information). Most of the *C. kullenbergi* isotopic data have been presented previously¹². The large isotopic offsets between *H. elegans* and *C. kullenbergi* are due to their different mineralogies. Benthic isotopes were measured on single shells, whereas planktonic $\delta^{18}\text{O}$ and benthic Cd_w used multiple (~3–15) shells per analysis. Of the 79 Cd measurements, three were considered to be significantly contaminated (Cd_w much higher than expected) and were omitted.

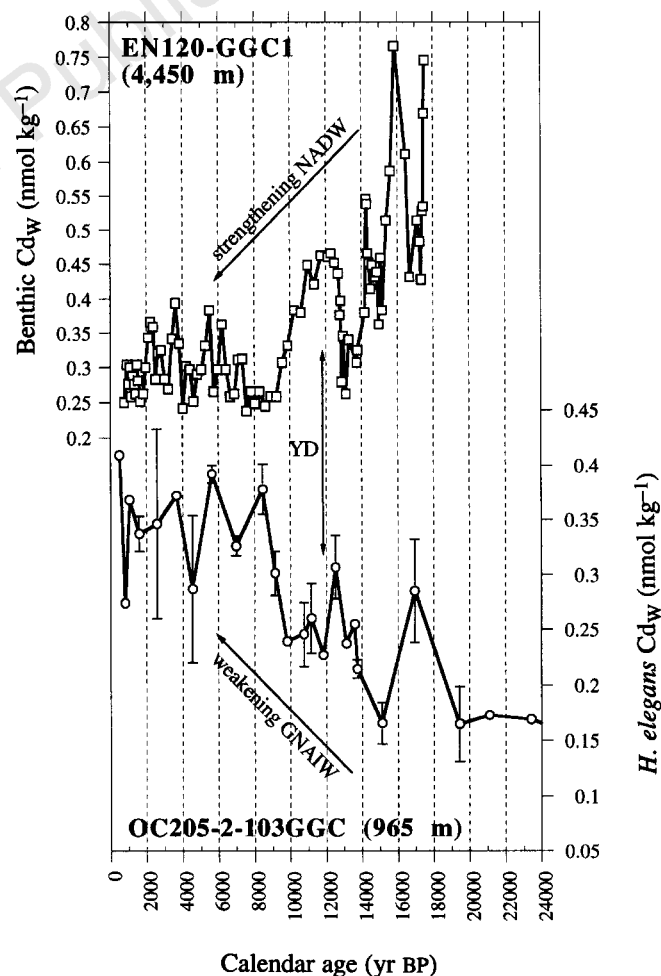


Figure 3 Records of Cd_w from the deep and intermediate-depth western North Atlantic. The EN120-GGC1 record (33° 40' N, 57° 37' W; 4,450 m; squares)^{3,9} is based on measurements of the benthic foraminifera *C. wuellerstorfi*, *Nutallides umbonifera*, and *Uvigerina* spp., using a D_p of 2.9. The OC205-2-103GGC record (26° 04' N, 78° 03' W; 965 m; circles) shows *H. elegans* means with standard errors. Dates for GGC1 are based on a multi-property correlation⁹ to nearby core KNR31-GPC5 (33° 41' N, 57° 37' W; 4,583 m), which has 13 AMS radiocarbon dates. Ten AMS dates were obtained for 103GGC. All radiocarbon dates were converted to calendar age BP using dendrochronological and coral-derived models (see Supplementary Information). 'YD' denotes the Younger Dryas.

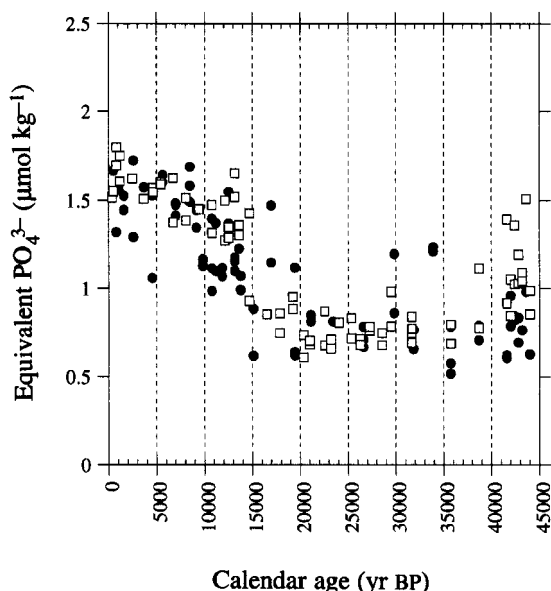


Figure 4 Estimated seawater PO_4^{3-} concentrations recorded in core OC205-2-103GGC. Estimates are based on *H. elegans* Cd_W (filled circles) and *C. kullenbergi* $\delta^{13}\text{C}$ (open squares). Cd_W was converted using the global Cd:P relationship: $\text{PO}_4^{3-} = 4.81 \text{ Cd}_W$ for $\text{Cd}_W < 0.28 \text{ nmol kg}^{-1}$, and $\text{PO}_4^{3-} = 0.638 + 2.51 \text{ Cd}_W$ for $\text{Cd}_W > 0.28 \text{ nmol kg}^{-1}$ (refs 16, 18). Although it has been suggested that the LGM ocean contained perhaps 13% less dissolved Cd than today, such a decrease is not well enough constrained to warrant the use of different equations for the glacial data^{20,27}. For data $\leq 5,000 \text{ yr BP}$, $\delta^{13}\text{C}$ was converted using the modern equation, $\text{PO}_4^{3-} = 2.45 - [0.91(\delta^{13}\text{C} - \delta^{13}\text{C}_{\text{as}})]$ (ref. 27). The glacial equation $\text{PO}_4^{3-} = 2.16 - [1.05(\delta^{13}\text{C} - \delta^{13}\text{C}_{\text{as}})]$ (ref. 27), which accounts for the glacial whole-ocean isotopic shift^{4,5}, was used for data $\geq 20,000 \text{ yr BP}$. Intermediate equations that evolve linearly with time were used between 5,000 and 20,000 yr BP. Based on $\delta^{13}\text{C}$ and Cd_W from the cores presented in Fig. 1, $\delta^{13}\text{C}_{\text{as}}$ was taken to be 0.25‰. This $\delta^{13}\text{C}_{\text{as}}$ may also contain some foraminiferal disequilibrium ('vital') effects.

therefore present what is, to our knowledge, the first direct palaeo-nutrient evidence for GNAIW formation during the Younger Dryas, at the expense of NADW.

The raw benthic $\delta^{13}\text{C}$ record from 103GGC looks significantly different from the Cd record (Fig. 2). Converting all $\delta^{13}\text{C}$ and Cd_W to equivalent PO_4^{3-} (refs 16, 18, 27), and accounting for the whole-ocean glacial carbon isotopic shift of 0.3‰ (refs 4, 5), again produces reasonably good agreement during the Late Holocene and LGM (Fig. 4). Although it is difficult to estimate how the whole-ocean $\delta^{13}\text{C}$ shift evolved across the deglaciation, a gradual linear change removes the $\delta^{13}\text{C}$ minimum seen at $\sim 13,000 \text{ cal. yr BP}$, improving the Cd- $\delta^{13}\text{C}$ agreement. Discrepancies remain, however, during the Younger Dryas and during stage 3 (before 40,000 cal. yr BP). These may be attributed to the obscuring effects of the whole-ocean isotopic shift and to the influences of air-sea exchange on $\delta^{13}\text{C}$ (ref. 19). For example, the opposing air-sea isotopic effects of cooling (raising $\delta^{13}\text{C}$) and decreased air-sea contact time (lowering $\delta^{13}\text{C}$) may have been roughly balanced during the LGM, but decreased air-sea contact may have dominated over the less extreme coolings of the Younger Dryas and stage 3. Given such complications to the $\delta^{13}\text{C}$ record, we believe that Cd is a more reliable indicator of short-term nutrient changes, especially within the main climate transitions.

Our results demonstrate that periods of enhanced intermediate-water production alternate with periods of enhanced deep-water formation on both orbital and millennial timescales. Analogous dynamics operate in the modern North Atlantic on much shorter (decadal) timescales, with deep convection alternating between

regions of upper (Labrador Sea) and lower (Greenland Sea) NADW formation²⁸. However, these brief reconfigurations have smaller effects on oceanic temperature and nutrient structure than the larger-scale events occurring during the last deglaciation.

The cooling of the North Atlantic during the Younger Dryas has long been linked to a reduction or cessation of NADW formation, which today releases a great deal of heat to the high-latitude atmosphere^{3,7-9}. Recent modelling suggests that a peak in atmospheric radiocarbon activity during the Younger Dryas is best explained by a cessation of NADW and a subsequent increase in GNAIW formation²⁹. Our Cd data from 965 m indeed indicate that GNAIW partially replaced NADW during this cold period, though to a lesser extent than during the LGM. A Younger Dryas $\delta^{13}\text{C}$ section through the eastern Atlantic⁶ supports this view of a moderately weakened NADW, but there are too few well-positioned shallow data in that reconstruction to discern a clear GNAIW mass; GNAIW should also be less evident in the eastern North Atlantic because deep waters are concentrated into western boundary currents. The Younger Dryas cooling may have thus been accomplished through the formation of intermediate waters that were less efficient at heating the North Atlantic than NADW³⁰. It is clear that North Atlantic climate change at the end of the last glacial period can no longer be explained by a simple convection 'on/off' switch. □

Received 22 July 1997; accepted 1 April 1998.

- Boyle, E. A. & Keigwin, L. D. Deep circulation of the North Atlantic over the last 200,000 years: Geochemical evidence. *Science* **218**, 784–787 (1982).
- Curry, W. B. & Lohmann, G. P. Reduced advection into Atlantic Ocean deep eastern basins during last glaciation maximum. *Nature* **306**, 577–580 (1983).
- Boyle, E. A. & Keigwin, L. D. North Atlantic thermohaline circulation during the past 20,000 years linked to high-latitude surface temperature. *Nature* **330**, 35–40 (1987).
- Curry, W. B., Duplessy, J. C., Labeyrie, L. D. & Shackleton, N. J. Changes in the distribution of $\delta^{13}\text{C}$ of deep water ΣCO_2 between the last glaciation and the Holocene. *Paleoceanography* **3**, 317–341 (1988).
- Duplessy, J. C. *et al.* Deepwater source variations during the last climatic cycle and their impact on the global deepwater circulation. *Paleoceanography* **3**, 343–360 (1988).
- Sarnthein, M. *et al.* Changes in east Atlantic deepwater circulation over the last 30,000 years: Eight time slice reconstructions. *Paleoceanography* **9**, 209–267 (1994).
- Broecker, W. S., Peteet, D. M. & Rind, D. Does the ocean-atmosphere system have more than one stable mode of operation? *Nature* **315**, 21–25 (1985).
- Broecker, W. S. *et al.* The chronology of the last deglaciation: Implications to the cause of the Younger Dryas event. *Paleoceanography* **3**, 1–19 (1988).
- Keigwin, L. D., Jones, G. A., Lehman, S. J. & Boyle, E. A. Deglacial meltwater discharge, North Atlantic deep circulation, and abrupt climate change. *J. Geophys. Res.* **96**, 16811–16826 (1991).
- Smith, J. E., Risk, M. J., Schwarz, H. P. & McConnaughey, T. A. Rapid climate change in the North Atlantic during the Younger Dryas recorded by deep-sea corals. *Nature* **386**, 818–820 (1997).
- Slowey, N. C. & Curry, W. B. Enhanced ventilation of the North Atlantic subtropical gyre thermocline during the last glaciation. *Nature* **358**, 665–668 (1992).
- Slowey, N. C. & Curry, W. B. Glacial-interglacial differences in circulation and carbon cycling within the upper western North Atlantic. *Paleoceanography* **10**, 715–732 (1995).
- Haddad, G. A. & Drozler, A. W. Metastable CaCO_3 dissolution at intermediate water depths of the Caribbean and western North Atlantic: Implications for intermediate water circulation during the past 200,000 years. *Paleoceanography* **11**, 701–716 (1996).
- Pedlosky, J. The dynamics of the oceanic subtropical gyres. *Science* **248**, 316–322 (1990).
- Olson, D. B., Schott, F. A., Zantopp, R. J. & Leaman, K. D. The mean circulation east of the Bahamas as determined from a recent measurement program and historical XBT data. *J. Phys. Oceanogr.* **14**, 1470–1487 (1984).
- Boyle, E. A. Cadmium: Chemical tracer of deepwater paleoceanography. *Paleoceanography* **3**, 471–489 (1988).
- Hester, K. & Boyle, E. A. Water chemistry control of the Cd content of benthic foraminifera. *Nature* **298**, 260–261 (1982).
- Boyle, E. A., Labeyrie, L. & Duplessy, J. C. Calcitic foraminiferal data confirmed by cadmium in aragonitic *Hoeglundina*: Application to the last glacial maximum in the northern Indian Ocean. *Paleoceanography* **10**, 881–900 (1995).
- Charles, C. D., Wright, J. D. & Fairbanks, R. G. Thermodynamic influences on the marine carbon isotope record. *Paleoceanography* **8**, 691–697 (1993).
- Boyle, E. A. Cadmium and $\delta^{13}\text{C}$ paleochemical ocean distributions during the Stage 2 glacial maximum. *Annu. Rev. Earth Planet. Sci.* **20**, 245–287 (1992).
- Fairbanks, R. G. A 17,000-year glacio-eustatic sea level record: influence of glacial melting rates on the Younger Dryas event and deep-ocean circulation. *Nature* **342**, 637–642 (1989).
- Zahn, R. & Keir, R. in *Carbon Cycling in the Glacial Ocean: Constraints on the Ocean's Role in Global Change* (eds Zahn, R. *et al.*) 195–221 (NATO ASI Ser. Vol. 117, Berlin, 1994).
- Wüst, G. On the stratification and the circulation in the cold water sphere of the Antillean-Caribbean basins. *Deep-Sea Res.* **10**, 165–187 (1963).
- CLIMAP Project Members. The surface of the ice-age Earth. *Science* **191**, 1131–1137 (1976).
- Boyle, E. A. The role of vertical chemical fractionation in controlling Late Quaternary atmospheric carbon dioxide. *J. Geophys. Res.* **93**, 15701–15714 (1988).
- Zahn, R. *et al.* Thermohaline instability in the North Atlantic during meltwater events: Stable isotope and ice-rafted detritus records from core SO75-26KL, Portuguese margin. *Paleoceanography* **12**, 696–710 (1997).
- Lynch-Stieglitz, J., van Geen, A. & Fairbanks, R. G. Inter-ocean exchange of Glacial North Atlantic Intermediate Water: Evidence from Subantarctic Cd/Ca and carbon isotope measurements. *Paleoceanography* **11**, 191–201 (1996).

28. Dickson, R., Lazier, J., Meincke, J., Rhines, P. & Swift, J. Long-term coordinated changes in the convective activity of the North Atlantic. *Prog. Oceanogr.* **38**, 241–295 (1996).
29. Hughen, K. A. *et al.* Deglacial changes in oceanic circulation from an extended radiocarbon calibration. *Nature* **391**, 65–68 (1998).
30. Rahmstorf, S. Rapid climate transitions in a coupled ocean-atmosphere model. *Nature* **372**, 82–85 (1994).

Supplementary Information is available on Nature's World-Wide Web site (<http://www.nature.com>) or as paper copy from the London editorial office of Nature.

Acknowledgements. We thank P. Lohmann for providing modern Bahamas PO_4^{3-} data and sediment grab samples; M. Jeglinski, D. Ostermann and L. Zou for isotope laboratory assistance; and E. Boyle, L. Keigwin, L. Labeyrie, D. McCorkle, J. McManus and R. Zahn for comments and discussions. This work was supported by the US NSF.

Correspondence and requests for materials should be addressed to T.M.M. (e-mail: tmarchitto@whoi.edu). Stable-isotope and Cd data are available on NOAA's World Data Center-A for Paleoclimatology (<http://www.ngdc.noaa.gov/paleo/paleodat.html>).

Iron-limited diatom growth and Si:N uptake ratios in a coastal upwelling regime

David A. Hutchins* & Kenneth W. Bruland†

* College of Marine Studies, University of Delaware, Lewes, Delaware 19958, USA

† Institute of Marine Sciences, University of California, Santa Cruz, Santa Cruz, California 95064, USA

There is compelling evidence that phytoplankton growth is limited by iron availability in the subarctic Pacific¹, and equatorial Pacific² and Southern oceans³. A lack of iron prevents the complete biological utilization of the ambient nitrate and influences phytoplankton species composition in these open-ocean 'high-nitrate, low-chlorophyll' (HNLC) regimes⁴. But the effects of iron availability on coastal primary productivity and nutrient biogeochemistry are unknown. Here we present the results of shipboard seawater incubation experiments which demonstrate that phytoplankton are iron-limited in parts of the California coastal upwelling region. As in offshore HNLC regimes, the addition of iron to these nearshore HNLC waters promotes blooms of large chain-forming diatoms. The silicic acid:nitrate (Si:N) uptake ratios in control incubations are two to three times higher than those in iron incubations. Diatoms stressed by a lack of iron should therefore deplete surface waters of silicic acid before nitrate, leading to a secondary silicic acid limitation of the phytoplankton community. Higher Si:cell, Si:C and Si:pigment ratios in diatoms in the control incubations suggest that iron limitation leads to more silicified, faster-sinking diatom biomass.

These results raise fundamental questions about the nature of nutrient-limitation interactions in marine ecosystems, palaeo-productivity estimates based on the sedimentary accumulation of biogenic opal, and the controls on carbon export from some of the world's most productive surface waters.

Our experiments used sea water collected from nearshore stations in the central California coastal upwelling regime. Subnanomolar dissolved Fe concentrations in the Big Sur area (Table 1) approach the low values in the remote oceanic HNLC regimes⁵. This region has little fluvial input and a narrow continental shelf, resulting in much smaller resuspended sedimentary Fe fluxes to surface waters than in many nearshore areas (K.W.B., unpublished observations). Our experiments suggest that during the upwelling season, abundant major nutrient supplies combine with low Fe inputs to make the Big Sur region a true coastal HNLC upwelling regime.

The results of one shipboard Fe-enrichment experiment (Big Sur 1997A) are presented in Fig. 1. Initial NO_3^- (12.3 μM), H_2SiO_3 (15.1 μM) and chlorophyll *a* (Chl *a*, 0.52 $\mu\text{g l}^{-1}$) concentrations were similar to typical open-ocean HNLC values. The picoplankton-dominated assemblage also resembled those of oceanic HNLC regions⁶, with only 23% of the total Chl *a* in the large ($>8 \mu\text{m}$) size fraction.

As in previous open-ocean Fe addition experiments^{4,7}, after a 2-day lag Chl *a* accumulated markedly in the Fe-enriched incubations, but only slightly in the controls (Fig. 1a). Separate experiments at a nearby station and at the same station gave similar results (Table 1, Big Sur 1996, Table 2, Big Sur 1997B). Net specific growth rates (μ) calculated from Chl *a* levels in these three experiments were 1.3 to 2.3 times higher in Fe-enriched incubations than in controls. The small increases in control Chl *a* concentration (Fig. 1a) are similar to those usually observed and often attributed to grazer exclusion in oceanic HNLC Fe-addition experiments³.

In a fourth experiment at Point Año Nuevo, north of Monterey Bay (37.05°N, 122.30°W), Chl *a* production and nutrient draw-down were unaffected by addition of Fe (Table 1). At Año Nuevo a broad, shallow shelf and nearby rivers supply ample Fe, and dissolved levels were much higher (1.5–2.6 nM) than at the Big Sur stations. Consequently, the phytoplankton community in this region was not Fe-stressed.

In oceanic Fe addition experiments, controls are usually dominated by nanoplankton and picoplankton, whereas diatoms bloom in Fe-enriched incubations^{1,4,7}. In our Big Sur experiments, large cells dominated in both treatments. By the end of the Big Sur 1997A and 1997B incubations, 83–90% (Fe-enriched) and 63–82% (controls) of the total Chl *a* was in the $>8 \mu\text{m}$ fraction (Tables 1 and 2). Most of this growth was by large diatoms, although large *Gymnodinium*-type dinoflagellates ($\sim 25 \mu\text{m}$) were also numerous.

Table 1 California coastal upwelling regime Fe addition experiments

Location and year Ambient [Fe]* (nM)	Treatment	Final Chl <i>a</i> † ($\mu\text{g l}^{-1}$)		Chl <i>a</i> net specific growth rate, μ (d ⁻¹)		Si:Chl <i>a</i> ‡ (mol:g)		Si:C§ (mol:mol)	Si:N (mol:mol)
		Total	$>8.0 \mu\text{m}$	Total	$>8.0 \mu\text{m}$	Total	$>8.0 \mu\text{m}$		
Big Sur 1997A 0.1–0.5	Control	1.9 ± 0.08	1.2 ± 0.03	0.30 ± 0.01	0.56 ± 0.01	5.8	9.6	0.27	1.6
	2.5 nM Fe	8.3 ± 0.59	6.9 ± 0.98	0.68 ± 0.02	1.01 ± 0.03	1.3	1.5	0.21	0.8
Big Sur 1996 0.06–0.3	Control	3.8 ± 2.5	5.0 ± 0.35	0.70 ± 0.07	0.56 ± 0.02	2.5	3.1	0.39	2.7
	10 nM Fe	9.2 ± 1.71	8.6 ± 0.26	0.94 ± 0.01	0.71 ± 0.01	1.5	1.8	0.24	1.1
Año Nuevo 1995 1.5–2.6	Control	13.2 ± 3.68	9.7 ± 3.20	0.44 ± 0.14	0.40 ± 0.16	1.0	1.5	n.d.	1.0
	10 nM Fe	10.8 ± 2.45	8.3 ± 1.90	0.37 ± 0.05	0.36 ± 0.14	1.1	1.4	n.d.	0.9

Errors are the standard deviations of three (Big Sur 1997A) or four (Año Nuevo 1995) replicate incubations, or the range of duplicate incubations (Big Sur 1996). All data are from the final time point (day 4) of each experiment, except for the Big Sur 1996 total Chl *a*, total growth rates and total Si:Chl *a* values. Final total Chl *a* samples from the day 4 time point for this experiment were lost, so reported data for this size class are from day 3 samples. n.d., no data.

* Fe concentrations are presented as a range of estimates. Lower estimates are reactive Fe analysed by cathodic stripping voltammetry of ultraviolet-oxidized samples²³. Upper estimates are total Fe determined by organic extraction of acidified samples, followed by graphite-furnace atomic absorption spectroscopy²⁴.

† Total Chl *a* was measured with GF/F glass-fibre filters (Big Sur 1996, Año Nuevo 1995) or 0.4 μm polycarbonate filters (Big Sur 1997A). $>8.0 \mu\text{m}$ Chl *a* was measured with polycarbonate filters.

‡ Silicic acid used:Chl *a* produced.

§ Silicic acid used:particulate organic carbon produced.

|| Silicic acid used:nitrate used.

But dinoflagellates did not respond to Fe addition; they were about equally abundant in both treatments (Table 2). In controls, final numbers of dinoflagellates and diatoms were similar, but in Fe-enriched incubations diatoms outnumbered dinoflagellates by > 3:1.

During the 4-day Big Sur 1997B incubation, numbers of all diatom species increased more than 20-fold in Fe-enriched incubations but only ~5-fold in controls (Table 2). Net specific growth rates of diatom cells were 0.76 d^{-1} (Fe-enriched) and 0.40 d^{-1} (controls). Final concentrations of the diatom-specific pigment fucoxanthin were about three times higher in Fe-enriched incubations than in controls (Table 2). Increases in all other taxon-specific pigments were small by comparison.

In Big Sur 1997B, the most common diatoms in the sparse initial assemblage were large centrics of the family Coscinodiscaceae

(Table 2). After 4 days the large chain-forming centric *Chaetoceros* dominated in both treatments, but was nearly six times more abundant in Fe-enriched incubations than in controls (Table 2). Net specific growth rates of *Chaetoceros* cells were 1.03 d^{-1} with Fe addition and 0.61 d^{-1} in controls. The chain-forming pennate diatom *Pseudonitzschia* also increased about sixfold in Fe-enriched incubations relative to controls. At the end of the incubation Coscinodiscaceae and cyanobacteria (*Synechococcus* spp.) were not dominant, but were more numerous in controls than in Fe-enriched incubations. Small unicellular pennate diatoms (*Navicula* spp.) and other eukaryotic nanoplankton were about equally abundant in both treatments (Table 2).

As in open ocean HNLC experiments, NO_3^- drawdown in the Big Sur 1997A incubation was strongly affected by the Fe additions (Fig. 1b): 99% of the original NO_3^- was depleted in the Fe-enriched incubations by day 4, whereas only 33% was used in the controls. These results were confirmed by the other two Big Sur experiments (Tables 1 and 2). Particulate organic nitrogen production⁸ closely balanced NO_3^- utilization, suggesting that almost all of the growth in both treatments was supported by nitrate-N.

Despite large differences in diatom biomass, H_2SiO_3 drawdown in Big Sur 1997A was only slightly greater in Fe-enriched incubations than in controls (Fig. 1c). In all three Big Sur experiments differential NO_3^- drawdown but similar H_2SiO_3 utilization resulted in different Si:N net uptake ratios between the two treatments. The molar ratio of Si:N used in the Fe-enriched incubations was always close to 1.0, typical of rapidly growing diatoms under nutrient-replete conditions⁹. However, Si:N uptake ratios were in every case two to three times higher in Fe-limited Big Sur controls (Tables 1 and 2). These estimates of Fe limitation effects on diatom Si:N uptake ratios are conservative, because other algal taxa might also assimilate NO_3^- . In the Fe-replete Año Nuevo experiment, Si:N drawdown ratios in both treatments were similar to the Big Sur Fe-enriched values (~1; Table 1).

Fe limitation can result not only in higher Si:N uptake ratios but also in more heavily silicified diatom biomass. Ratios of Si used to Chl *a* produced in the Big Sur experiments were 1.7 to 6.4 times higher in controls than in Fe-enriched incubations (Tables 1 and 2). In the high-Fe Año Nuevo experiment, Si:Chl *a* ratios in both treatments were similar to the Big Sur Fe-enriched values. Si:Chl *a* ratios depend partly on cell Chl *a* content, which is reduced during Fe-limited growth¹⁰. However, other evidence also indicates that Fe-stressed diatom communities are more highly silicified.

Si used per diatom cell produced (mol per cell) in the Big Sur 1997B incubation was 2.9 times higher in controls than in Fe-enriched incubations. Ratios of Si used to fucoxanthin produced were 1.7 times greater in controls (Table 2). In the other Big Sur experiments, molar ratios of Si used to particulate organic carbon produced⁸ (Si:C) were 1.3 to 1.6 times higher in controls (Table 1). Martin's data indicate that Si:C ratios are often ~0.4 in oceanic HNLC experimental controls but are ~0.2 after Fe addition¹¹. Particulate Si:C ratios are minimum estimates of diatom ratios, owing to carbon in other cells and detritus. Diatom opal is ~50% denser (~2.2 g cm^{-3}) than proteins or carbohydrates (~1.5 g cm^{-3})¹², suggesting that diatoms in Fe-limited areas are heavier and will sink faster.

Control of diatom silicification and Si:N uptake ratios by Fe has global biogeochemical implications. Dugdale and co-workers^{13,14} suggest that in regimes dominated by diatoms, grazers preferentially export Si relative to more efficiently recycled N, driving the system towards Si limitation. Once Si has been depleted, further diatom growth is prevented even when other nutrients (N, P, Fe, etc.) are available. The 'silicate pump' thus favours a picoplankton-dominated community that is growing largely on recycled nutrients and that exports carbon inefficiently. Our results suggest that Fe limitation needs to be considered as an important force driving the silicate pump.

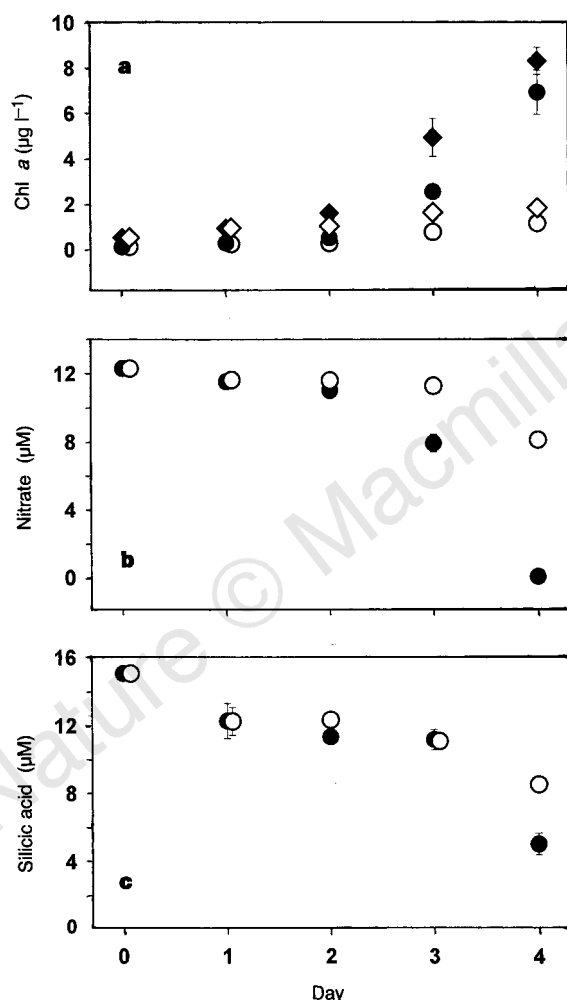


Figure 1 Chl *a* and nutrient concentrations in the Big Sur 1997A Fe addition experiment. Near-surface water (5–10 m) was sampled on 22 June 1997 at 35°58' N, 122°01' W with a trace-metal-clean Teflon pumping system. Experimental bottles (+2.5 nM FeCl_2 in 0.1 M HCl (Ultrax) and unamended controls) were sampled daily during a 4-d incubation in a flow-through, spectrally corrected deckboard incubator¹⁹. Error bars (standard deviations of triplicate 2.7-litre polycarbonate bottles) are included on all data points, but in many cases are smaller than the symbols. **a**, Size-fractionated Chl *a* in the +Fe treatments (◆, >0.2 µm; ●, >8.0 µm) and the controls (◇, >0.2 µm; ○, >8.0 µm). Chl *a* was measured fluorometrically on samples vacuum-filtered on to polycarbonate filters and extracted in 90% acetone¹⁹. **b**, Nitrate concentrations in the Fe incubations (●) and controls (○). **c**, Silicic acid concentrations in the two treatments; symbols as in **b**. Preliminary nutrient concentrations were monitored at sea with an autoanalyser and were confirmed and corrected using frozen samples in the laboratory⁷.

Table 2 Big Sur 1997B Fe-addition experiment

Biological and chemical variables	Initial	+Fe final	Control final
>0.2 µm Chl <i>a</i> (µg l ⁻¹)	0.54 ± 0.07	9.20 ± 0.80	2.30 ± 0.25
Net specific growth rate (d ⁻¹)		0.71 ± 0.02	0.36 ± 0.03
>8.0 µm Chl <i>a</i> (µg l ⁻¹)	0.12 ± 0.02	8.30 ± 0.64	1.90 ± 0.17
Net specific growth rate (d ⁻¹)		1.06 ± 0.02	0.69 ± 0.03
Fucoxanthin (µg l ⁻¹)	0.07 ± 0.01	3.55 ± 0.35	1.24 ± 0.23
Net specific growth rate (d ⁻¹)		0.98 ± 0.03	0.71 ± 0.05
Diatoms (cells l ⁻¹)			
<i>Chaetoceros</i>	(0.05 ± 0.05) × 10 ⁶	(2.95 ± 1.30) × 10 ⁶	(0.54 ± 0.35) × 10 ⁶
<i>Pseudonitzschia</i>	(0.04 ± 0.01) × 10 ⁶	(0.74 ± 0.35) × 10 ⁶	(0.12 ± 0) × 10 ⁶
<i>Coscinodiscaceae</i>	(0.09 ± 0.03) × 10 ⁶	(0.07 ± 0.02) × 10 ⁶	(0.15 ± 0.09) × 10 ⁶
<i>Navicula</i>	(0.01 ± 0) × 10 ⁶	(0.18 ± 0.07) × 10 ⁶	(0.12 ± 0.04) × 10 ⁶
All diatom species	0.19 × 10 ⁶	3.93 × 10 ⁶	0.94 × 10 ⁶
Dinoflagellates (all species cells l ⁻¹)	(0.15 ± 0.1) × 10 ⁶	(1.13 ± 0.11) × 10 ⁶	(1.25 ± 0.06) × 10 ⁶
Eukaryotic nanoplankton (all species cells l ⁻¹)	(0.11 ± 0.03) × 10 ⁶	(0.60 ± 0.31) × 10 ⁶	(0.54 ± 0.45) × 10 ⁶
Cyanobacteria (all species cells l ⁻¹)	n.o.	(0.22 ± 0.38) × 10 ⁶	(0.48 ± 0.18) × 10 ⁶
Nitrate (µM)	8.1 ± 0	0.13 ± 0.23	6.5 ± 0.20
Silicic acid (µM)	8.2 ± 0	0 ± 0	3.4 ± 0.24
Si:N uptake (mol:mol)		1.0	3.0
Si used:>0.2 µm Chl <i>a</i> produced (mol:g)		0.9	2.7
Si used: >8.0 µm Chl <i>a</i> produced (mol:g)		1.0	2.7
Si used: fucoxanthin produced (mol:g)		2.4	4.1
Si used: diatom cells produced (mol:cell)		2.2 × 10 ⁻¹²	6.4 × 10 ⁻¹²

Species composition was measured with epifluorescence microscopy cell counts and high-performance liquid chromatography measurements of phytoplankton taxon-specific pigments⁷. Errors are the standard deviations of triplicate incubations. n.o., not observed.

Effects of Fe on the silicate pump are probably important not only in coastal HNLC areas but also in the three large oceanic HNLC regimes. Unfortunately, many previous Fe addition experiments in these areas omitted Si measurements, or data sets are incomplete and anecdotal.

Examination of available field data shows that in all three Fe-limited oceanic HNLC regimes the H₂SiO₃:NO₃⁻ ratio of freshly upwelled water is close to or greater than 1, but as surface water advects away from the upwelling centre this ratio drops. As newly upwelled water from the centre of the subarctic Pacific HNLC area advects and mixes 7° southwards, Si:N drawdown occurs at a molar ratio of ~2.3 (ref. 15). Drawdown ratios are often over 4 in Fe-limited areas of the Southern Ocean¹⁶. At the Antarctic Polar Front a slowly upwelled Fe flux of ~0.05 nM d⁻¹ into the upper 100 m supported a drawdown of 10 µM H₂SiO₃ but only 1–2 µM NO₃⁻ over 3 weeks, a Si:N uptake ratio of >5 (ref. 17).

Surface water transects (K.W.B., unpublished observations) demonstrate that Fe concentrations are correlated with Si:N drawdown ratios off California. Freshly upwelled waters here have nearly equimolar concentrations of NO₃⁻ and H₂SiO₃ (~12–25 µM). In aged surface waters in high-Fe areas (>1 nM), NO₃⁻ and H₂SiO₃ are generally both depleted to low concentrations, suggesting equimolar drawdown. In low-Fe waters such as along the Big Sur coast, H₂SiO₃ is often rapidly exhausted while NO₃⁻ remains high. High Si:N drawdown ratios that result in H₂SiO₃-depleted, NO₃⁻-replete surface waters promise to be a reliable diagnostic indicator of Fe-limited upwelled waters in this system, and possibly in others.

Our experiments suggest that Fe-stressed diatom assemblages will deplete Si from the water column before N, because H₂SiO₃ utilization is not regulated by Fe availability to the extent that NO₃⁻ uptake is. Some diatoms have elevated Si:N ratios when NH₄⁺ is limiting¹⁸, and Fe limitation might also increase Si:N ratios by decreasing N assimilation as NO₃⁻. Differential drawdown of H₂SiO₃ and NO₃⁻ in advecting HNLC water masses occurs slowly (days or weeks) as periodic small inputs and recycling of Fe¹⁹ allow marginal diatom growth. This effect might not be evident over the duration of a typical oceanic Fe addition experiment, but happens more quickly in coastal waters where diatoms are initially somewhat more abundant and not as severely Fe-limited.

Lower Si:N and Si:C ratios in our Fe-enriched incubations might be partly due to Fe-induced changes in floristics. Elemental ratios in diatoms are not correlated with surface:volume ratios⁹, so

this difference is probably not due to the greater abundance of large cells in Fe-amended samples. However, cultures of the genus (*Chaetoceros*) that bloomed in our Fe-enriched incubations have relatively low Si:C (~0.1) and Si:N (~1.0) ratios. Some unicellular centric diatoms, which were relatively more abundant in our controls, are more highly silicified (for example *Coscinodiscus*, Si:C = 0.2–0.4, Si:N ≈ 2.0)⁹.

Si per cell within a species might also be higher under Fe limitation. Si uptake normalized per diatom cell was much higher in our controls (Table 2). Although these data were necessarily calculated from numbers of all diatom species combined, *Chaetoceros* was numerically dominant in both controls and Fe-enriched incubations (58% and 75% of total diatom cells, respectively). Sinking rates of a cultured subarctic diatom are five times faster during Fe-limited growth²⁰, consistent with higher cellular Si:C ratios. Thus Fe limitation might produce more silicified individuals, select for more silicified species, or both. Regardless, faster sinking diatom biomass could result in proportionately more efficient carbon export in low-Fe regimes, despite lower total productivity.

These results have implications for the use of biogenic opal in sediments as a proxy for changes in limitation of productivity by Fe^{21,22}. If Fe-limited diatom assemblages have higher Si:C ratios, opal-derived calculations of diatom palaeoproductivity during times of decreased Fe inputs (such as interglacial periods) might be overestimates. Decreased dissolution of silica in the water column owing to faster sinking rates and preferred preservation of robust, heavily silicified diatom frustules in sediments could further increase this discrepancy.

Our experiments demonstrate that Fe limitation and HNLC conditions are not unique to remote oceanic upwelling areas. Local hydrological, geological and meteorological conditions can result in Fe limitation in parts of major coastal upwelling regions as well. The Peru upwelling in particular is noteworthy for the lack of an extensive continental shelf and minimal freshwater inflow, and is often characterized by HNLC conditions¹⁶ and very high Si:N drawdown ratios¹³. These areas have a disproportionate role in global marine new production, and potential Fe limitation of their diatom-dominated communities has large implications for carbon cycling in the ocean. Fe can control phytoplankton growth and species composition in both coastal and oceanic upwelling areas, and has a more subtle but none the less crucial role in

determining the balance between Si and N uptake by the biological community. □

Received 29 July 1997; accepted 21 February 1998.

- Martin, J. H. & Fitzwater, S. E. Iron deficiency limits phytoplankton growth in the north-east Pacific subarctic. *Nature* **331**, 341–343 (1988).
- Coale, K. H. *et al.* A massive phytoplankton bloom induced by an ecosystem-scale iron fertilization experiment in the equatorial Pacific Ocean. *Nature* **383**, 495–501 (1996).
- de Baar, H. J. W. *et al.* On iron limitation of the Southern Ocean: experimental observations in the Weddell and Scotia Seas. *Mar. Ecol. Prog. Ser.* **65**, 105–122 (1990).
- Hutchins, D. A. in *Progress in Phycological Research* Vol. II (eds Chapman, D. & Round, F.) 1–49 (Biopress, Bristol, 1995).
- Bruland, K. W., Donat, J. R. & Hutchins, D. A. Interactive influences of bioactive trace metals on biological production in oceanic waters. *Limnol. Oceanogr.* **36**, 1555–1577 (1991).
- Chavez, F. P. *et al.* Phytoplankton variability in the central and eastern tropical Pacific. *Deep-Sea Res. II* **43**, 835–870 (1996).
- DiTullio, G. R., Hutchins, D. A. & Bruland, K. W. Interaction of iron and major nutrients controls phytoplankton growth and species composition in the tropical north Pacific Ocean. *Limnol. Oceanogr.* **38**, 495–506 (1993).
- Sharp, J. H. Improved analysis for 'particulate' organic carbon and nitrogen from seawater. *Limnol. Oceanogr.* **19**, 984–989 (1974).
- Brzezinski, M. A. The Si:C:N ratio of marine diatoms: interspecific variability and the effect of some environmental variables. *J. Phycol.* **21**, 347–357 (1985).
- Geider, R. J. & LaRoche, J. The role of iron in phytoplankton photosynthesis, and the potential for iron-limitation of primary productivity in the sea. *Photosyn. Res.* **39**, 275–301 (1994).
- Martin, J. H. in *Primary Productivity and Biogeochemical Cycles in the Sea* (eds Falkowski, P. G. & Woodhead, A. D.) 123–137 (Plenum, New York, 1992).
- Handbook of Chemistry and Physics* 63rd edn (eds Weast, R. C. & Astle, M. J.) (CRC, Boca Raton, 1982).
- Dugdale, R. C., Wilkerson, F. P. & Minas, H. J. The role of a silicate pump in driving new production. *Deep-Sea Res.* **142**, 697–719 (1995).
- Dugdale, R. C. & Wilkerson, F. P. Silicate regulation of new production in the eastern equatorial Pacific. *Nature* **391**, 270–273 (1998).
- Martin, J. H., Gordon, R. M., Fitzwater, S. & Broenkow, W. W. VERTEX: phytoplankton/iron studies in the Gulf of Alaska. *Deep-Sea Res.* **36**, 649–680 (1989).
- Minas, H. J. & Minas, M. Net community production in 'High Nutrient-Low Chlorophyll' waters of the tropical and Antarctic Oceans: grazing versus iron hypothesis. *Ocean. Acta* **15**, 145–162 (1992).
- de Baar, H. J. W. *et al.* Importance of iron for plankton blooms and carbon dioxide drawdown in the Southern Ocean. *Nature* **373**, 412–415 (1995).
- Harrison, P. J., Conway, H. L., Holmes, R. W. & Davis, C. O. Marine diatoms grown in chemostats under silicate or ammonium limitation. III. Cellular composition and morphology of *Chaetoceros debilis*, *Skeletonema costatum* and *Thalassiosira gravida*. *Mar. Biol.* **43**, 19–31 (1977).
- Hutchins, D. A., DiTullio, G. R. & Bruland, K. W. Iron and regenerated production: evidence for biological iron recycling in two marine environments. *Limnol. Oceanogr.* **38**, 1242–1255 (1993).
- Muggli, D. L., Lecourt, M. & Harrison, P. J. Effects of iron and nitrogen source on the sinking rate, physiology and metal composition of an oceanic diatom from the subarctic Pacific. *Mar. Ecol. Prog. Ser.* **132**, 215–227 (1996).
- Mortlock, R. A. *et al.* Evidence for lower productivity in the Antarctic Ocean during the last glaciation. *Nature* **351**, 220–223 (1991).
- Charles, C. D., Froelich, P. N., Zibello, M. A., Mortlock, R. A. & Morley, J. J. Biogenic opal in Southern Ocean sediments over the last 450,000 years: implications for surface water chemistry and circulation. *Paleoceanography* **6**, 697–728 (1991).
- Rue, E. L. & Bruland, K. W. The role of organic complexation on ambient iron chemistry in the equatorial Pacific Ocean and the response of a mesoscale iron addition experiment. *Limnol. Oceanogr.* **42**, 901–910 (1997).
- Bruland, K. W., Franks, R. P., Knauer, G. A. & Martin, J. H. Sampling and analytical methods for the determination of copper, cadmium, zinc and nickel at the nanogram per liter level in sea water. *Anal. Chim. Acta* **105**, 233–245 (1979).

Acknowledgements. We thank M. Sanderson, C. De La Rocha, Y. Zhang, M. Schwartz, G. Smith, M. Brzezinski, V. Franck, N. Fisher, P. Harrison, D. Kirchman, S. Wilhelm, A. Witter and the captain and crew of the RV *Pt. Sur*. This work was supported by NSF Chemical and Biological Oceanography and University of Delaware Research Foundation funding.

Correspondence and requests for material should be addressed to D.A.H. (e-mail address: dahutch@udel.edu).

Anisotropic structures at the base of the Earth's mantle

Lev Vinnik^{†‡}, Ludovic Breger[‡] & Barbara Romanowicz^{*‡}

^{*} Seismological Laboratory and [‡] Department of Geology and Geophysics, University of California, Berkeley, California 94720, USA

[†] Institute of Physics of the Earth, Bolshaya Gruzinskaya 10, Moscow CIS

The D'' shell at the base of the Earth's mantle is thought to be a thermal and compositional boundary layer where vigorous dynamical processes are taking place^{1–4}. An important property of D'' is its seismic anisotropy, expressed as different velocities for horizontally and vertically polarized shear waves that have been diffracted or reflected at the core–mantle boundary^{5,6}. The nature of this anisotropy has been the subject of debate^{7–11}.

Here we present an analysis of various seismic phases, generated in the Kermadec–Fiji–Tonga zone and recorded at stations in North America, which reveal a region at the base of the mantle beneath the southwest Pacific Ocean where horizontally propagating vertically polarized waves are slower (by at least 10 per cent) than horizontally polarized waves. This observed anisotropy is an order of magnitude larger than that previously thought to exist in the lower mantle, and corresponds to lateral variations in horizontally polarized shear-wave velocity which are also of about 10 per cent. We speculate that this anisotropy may be the result of the mixing and shearing of strongly heterogeneous material in the boundary layer.

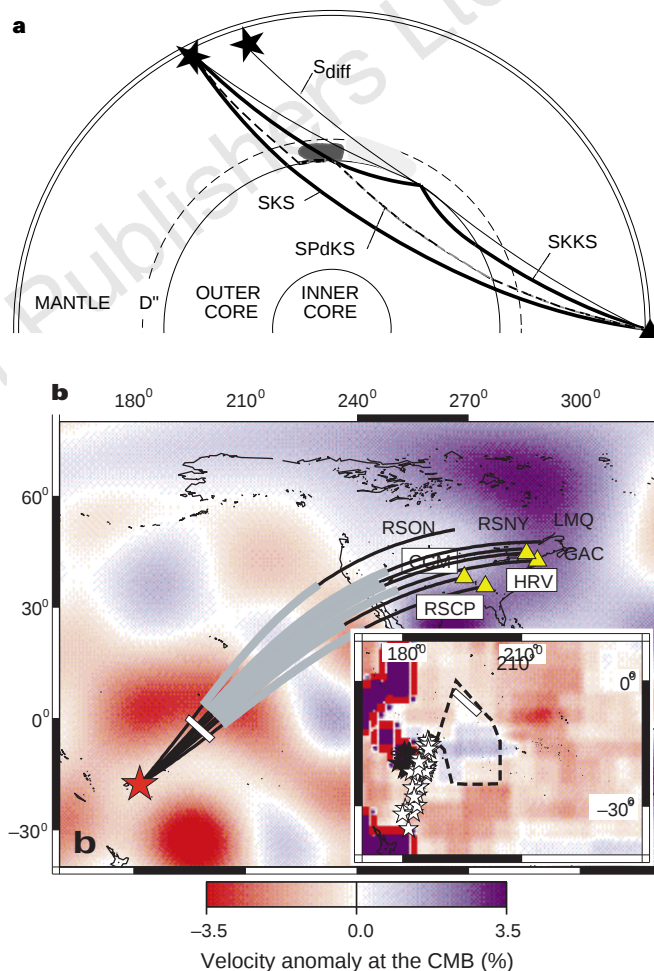


Figure 1 Geometry of the data set considered. **a**, Schematic representation of the wave paths of SKS, SKKS, Sdiff and SPdKS. The wave paths of Sdiff for two different positions of the source (stars) differ only at the source side. The lightly shaded area indicates the region where SH is slow, as discussed in the text. The darker, hatched area is the region where SH velocity is elevated or normal, but SV is much slower than SH. **b**, Surface projections of wave paths superimposed on a tomographic model of SH velocity in D'' (ref. 15). The light-grey zone of the wave path indicates approximately the D'' leg of Sdiff for the shortest distances considered. The white bar corresponds to the kink for SHdiff–SKS residuals, discussed in the text. Red star is location of earthquake source region; yellow triangles and 3/4-letter codes indicate station locations and names, respectively. Inset, close-up look at the source side of the wave-paths. Epicentres of deep Fiji–Tonga events and of shallow Kermadec–Fiji–Tonga events are shown by filled and open stars, respectively. The background tomographic model of S-wave velocity in D'' is from ref. 16. The polygon indicates the zone of ultra-low P-wave velocity in D'' (ref. 18). The barcode indicates amplitudes of shear velocity anomalies at the core–mantle boundary (CMB) in the tomographic models^{15,18}.

The first part of our analysis is based on the residuals of differential travel times between seismic phases SH_{diff} and SKS/SKKS relative to those for the standard Earth model PREM¹². (SH_{diff} are horizontally polarized S waves that have been diffracted at the core–mantle boundary; SKS/SKKS are shown in Fig. 1.) Differential travel times are helpful because they are not sensitive to errors in focal parameters of seismic events and structural complexities in the source region. Contrary to earlier studies^{13,14}, we consider (1) not the absolute values of the residuals, but rather the slopes of the resulting time–epicentral distance plots; (2) not the average residuals for many stations, but the individual slope for every station; and (3) not only the residuals for deep events, but also for intermediate and shallow events. For the given source–receiver geometry (Fig. 1) and a fixed station, changes of the wave path occur practically only at the source side of the wave path.

The residuals as a function of epicentral distance show a remarkable trend, consistent from station to station, for shallow and deep events and whether SKS or SKKS is used as reference phase. Figure 2 shows examples for the three best stations. Each panel can be divided into two parts, with rapidly increasing residuals at shorter distances, and slowly rising (or decreasing) residuals at larger distances. To make this trend more visible, we divide each set of measurements into two subsets corresponding to epicentral distances smaller than or equal to, and greater than or equal to, a variable cut-off distance. When the best-fitting position of the cut-off and linear fits on both sides of it are obtained by regression, the average variance reduction is twofold with respect to a single regression line. We note that the epicentral distance of the change of slope (the ‘kink’) varies (for example, 116–117° at HRV; 103–104° at RSON), depending on the epicentral distance range of the Kermadec–Fiji–Tonga seismic zone to the given station. Comparable results are obtained for the other stations where a similar analysis was performed: RSCP, LMQ, RSNY and GAC (Fig. 1).

The features in Fig. 2 cannot be explained by mantle complexity immediately beneath the seismic foci, because the change of slope is

present in the data for both deep and shallow events. Therefore, our preferred explanation for the kink is a first-order change in the properties of the lowermost mantle at the source side of the wave-path. Moreover, SKS and SKKS ray-paths are separated in D'' by a distance exceeding 10°, and a similar trend in the residuals in Fig. 2 with respect to SKS and SKKS implies that the effect is mainly in the travel times of SH_{diff} . If the position of the kink for every station is inverted for the position of the corresponding border in the D'' layer, assuming that SH enters D'' at a height of 300 km above the core–mantle boundary, the estimates for different stations are mutually consistent and correspond to the white bar in Fig. 1b. In experiments with synthetic seismograms, the anomaly of slowness of 1.1 s per degree corresponds to S-wave velocity in D'' reduced by ~10% with respect to PREM. The region of anomalously low SH-wave velocity is located to the northeast of the bar (Fig. 1b), whereas SH velocity to the southwest of the bar is either slightly elevated or normal. Qualitatively, this division is confirmed by two recent tomographic models^{15,16} (Fig. 1b), but the magnitude of the low-velocity anomaly in both models is ~3 times smaller than in our data. Robustness of our technique is confirmed by other data¹⁷.

The residuals of SKKS–SKS differential travel times relative to PREM, derived from our data (Fig. 3) are positive, with an average value of around 2–3 s. As argued in ref. 18, these positive residuals can only be explained by anomalously low S-wave velocity in the lowermost mantle on the source side, due to the longer paths of SKKS relative to SKS. Our data, however, indicate that, on the source side, SKS and SKKS propagate in a region of D'' with normal SH velocity (Fig. 1b). As SKS and SKKS are SV polarized, to explain the discrepancy at least partly, we suggest that although SH velocity is normal, SV velocity in the lowermost mantle on the source side is anomalously low. (SV indicates vertically polarized S waves.)

To obtain quantitative estimates of this anisotropy, we assume that the lowermost mantle is intrinsically isotropic, finely layered and horizontally stratified. For long waves, it behaves like a homogeneous transversely isotropic medium with a vertical axis of

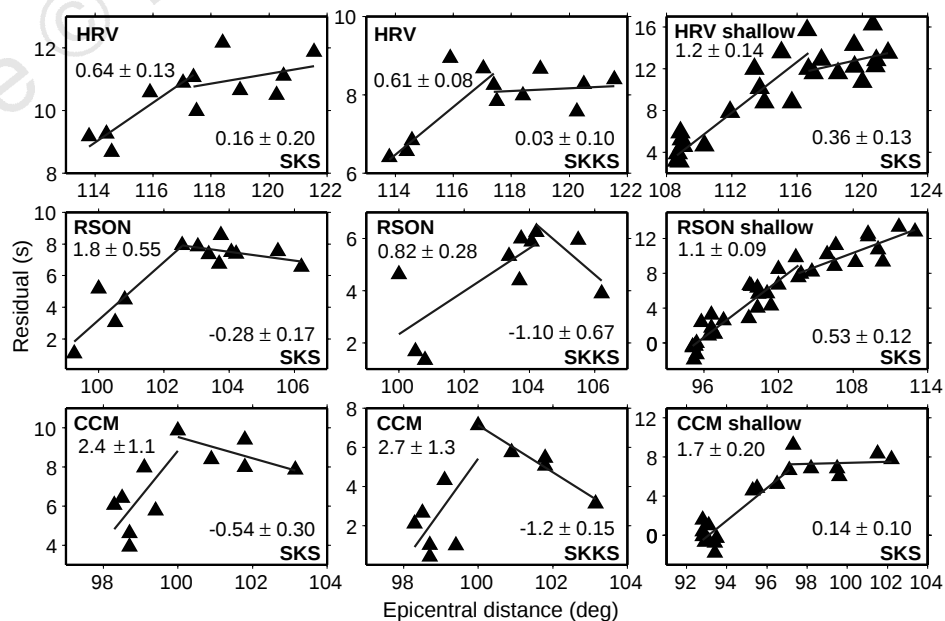


Figure 2 Residuals of SH_{diff} –SKS and SH_{diff} –SKKS differential travel times with respect to those for PREM¹², for HRV/WFM, RSON and CCM, and corresponding regression lines. The data labelled ‘shallow’ are for shallow and intermediate events; all others are for deep events. Shallow events broaden the distance interval available for SKS. The slopes of the plots are controlled mainly by the slowness of SH_{diff} and, consequently, by SH velocity at the source side of D'' . The slope is negative, positive or equal to zero, if SH velocity is higher, lower or similar

to that in PREM. Numerical values of the slopes of the regression lines with their respective confidence intervals are shown in the upper left and lower right corners of each panel. Confidence intervals are usually around ± 0.1 – 0.3 s per degree, but they exceed ± 1.0 s per degree for deep events at CCM, left of the kink. If the latter data are excluded, the average values of the slopes are 1.12 ± 0.2 and -0.21 ± 0.2 s per degree, left and right of the kink, respectively.

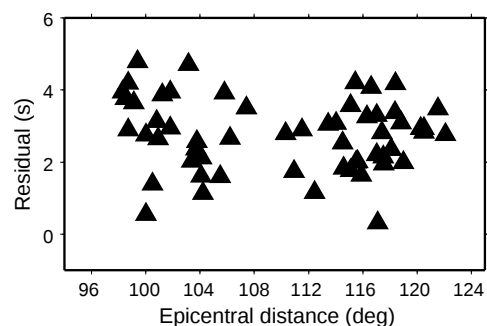


Figure 3 SKKS-SKS residuals for all paths in Fig. 1b. Comparable residuals were reported for similar wave-paths in earlier studies^{13,18,24}. For other paths beneath the Pacific and America, the residuals are close to zero^{18,24}.

symmetry^{19–21}. In this medium, SH_{diff} and SV_{diff} propagate independently, with the former faster than the latter, which is consistent with the seismic observations⁵. To explain residuals of ~ 2 s, SKS and SKKS velocities in a layer 300 km thick should be $\sim 10\%$ lower than standard velocity. For the angles of incidence characteristic of SKS and SKKS, the SV velocities are larger than for horizontal propagation. Then the difference between SH and SV velocities for horizontal propagation can be even higher than 10%. This prediction can be tested by observing propagation of SH_{diff} and SV_{diff} .

The seismic phase SV_{diff} is usually weak, and can be easily distorted by effects of azimuthal anisotropy in the mantle outside D'' and side refraction of SH_{diff} . To eliminate these effects and to pick arrivals of SV_{diff} accurately, we have devised a special technique²², illustrated in ref. 22 for the records of HRV/WFM and RSON. For the present study, a search for SV_{diff} has been conducted in the records of the deep Fiji-Tonga events at seismograph stations shown in Fig. 1b. The best data set has been obtained at HRV/WFM, RSON and CCM. Remarkable features of the detected SV_{diff} signals are their distance-dependent delays relative to SH_{diff} (Fig. 4). At shorter distances, the slopes of the plots are close to 0 s per degree, whereas at larger distances they are in the range 1.0–1.6 s per degree. Then anisotropy is weak (not more than a few per cent) in the region of very low SH velocity (northeast of the white bar in Fig. 1b), and $\sim 15\%$ in the region of normal SH velocity, southwest of the bar. Anisotropy of $\sim 15\%$ is stronger by about an order of magnitude than reported elsewhere.

Additional data on the properties of D'' beneath the southwest Pacific are provided by the observations of SPdKS (Fig. 1a). This phase propagates as P_{diff} in D'' and as SKS elsewhere²³. P_{diff} beneath the southwest Pacific is anomalously slow^{18,24}, with a remarkable correlation between the anomalous delays of SPdKS relative to SKS and positive SKKS-SKS residuals, similar to those shown in Fig. 3. An explanation for this phenomenon is partial melting²⁵. The anomalously slow P_{diff} , however, propagates on the source side in a medium with normal SH velocity (Fig. 1b), which is hard to reconcile with a strong reduction of S velocity, expected in the case of partial melting. Because, as we have demonstrated, the SKKS-SKS residuals can be related to anisotropy, the correlation between them and SPdKS delays suggests that the delays are affected by anisotropy as well. Moreover, east of the white bar in Fig. 1b, where (according to our data) anisotropy is weak, the layer of low P-wave velocity, if present, is very thin²⁶. Assuming that the SPdKS delays are related to anisotropy, we note that the low horizontal P-wave velocity in a fine-layered horizontally stratified medium is possible if the variations of S-wave velocity between the layers are much stronger than those of P-wave velocity²⁰. Another possibility to be considered is lattice preferred orientation²⁷.

For normal distribution of random variations of S-wave velocity in the stack of thin horizontal layers with mean m and standard

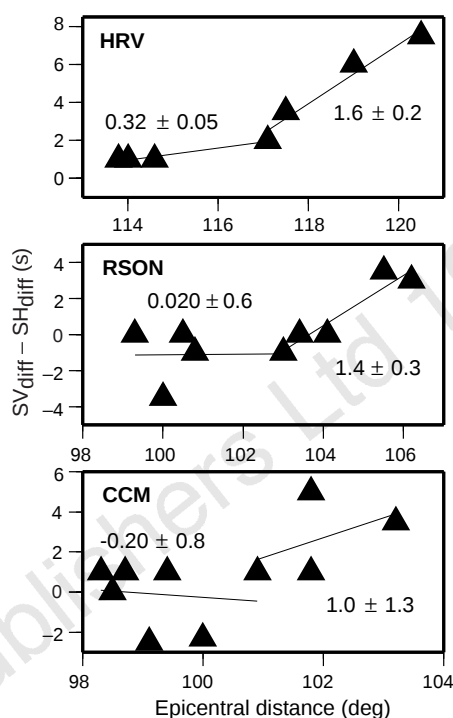


Figure 4 Delays of SV_{diff} with respect to SH_{diff} as a function of epicentral distance for stations HRV/WFM, RSON and CCM. The slope of any of the plots is the difference in slowness between SV_{diff} and SH_{diff} at the source side of the wave-path. Magnitude of anisotropy can be determined as the ratio between these slopes, and the standard slowness of 8.4 s per degree. Each data set is approximated by two regression lines. The boundaries between the corresponding distance intervals are taken from Fig. 2. Numerical values of the slopes of the regression lines with their respective confidence intervals are shown adjacent to the corresponding portions of the plots.

deviation σ , the SH/SV velocity ratio for horizontal propagation is expressed²⁰ as $SH/SV \approx \sqrt{1 + 4(\sigma/m)^2}$. For S-wave anisotropy of 15%, σ thus evaluated is 28% of m . It is not possible to explain the variations exceeding a few per cent solely by temperature variations²⁸. S-wave velocity can be significantly lowered by partial melting or/and by accumulation of crystalline iron-alloy products of chemical reactions between iron of the outer core and mantle perovskite²⁹. Among these possibilities, relatively weak variations of P-wave velocity favour partial melting. The layered structure in D'' could be generated by convective mixing and shearing. Then, strong anisotropy in D'' should be accompanied by strong wave scattering. The anomalous region of D'' , if projected on the surface of the Earth, is close to Polynesia, a region of unusual thermal agitation, where a large-scale thermochemical plume may be present at the top of the lower mantle³⁰. The anisotropic region in D'' could somehow be related to the same plume.

Comparable magnitudes of lateral heterogeneity and anisotropy suggest that lateral variations of anisotropy can be mistaken for lateral velocity variations, when the models are derived under the assumption of isotropy. Whatever its origin, the observations of strong and laterally variable anisotropy add a new dimension to the question of properties and processes in the D'' layer. □

Received 27 March 1997; accepted 19 March 1998.

1. Doornbos, D. J., Spiliopoulos, S. & Stacey, F. D. Seismological properties of D'' and the structure of a thermal boundary layer. *Phys. Earth Planet. Inter.* **41**, 225–239 (1986).
2. Knittle, E. & Jeanloz, R. Earth's core-mantle boundary: results of experiment at high pressures and temperatures. *Science* **251**, 1438–1443 (1991).
3. Kellogg, L. H. & King, S. D. Effect of mantle plumes on the growth of D'' by reaction between the core and mantle. *Geophys. Res. Lett.* **20**, 379–382 (1993).

- 567

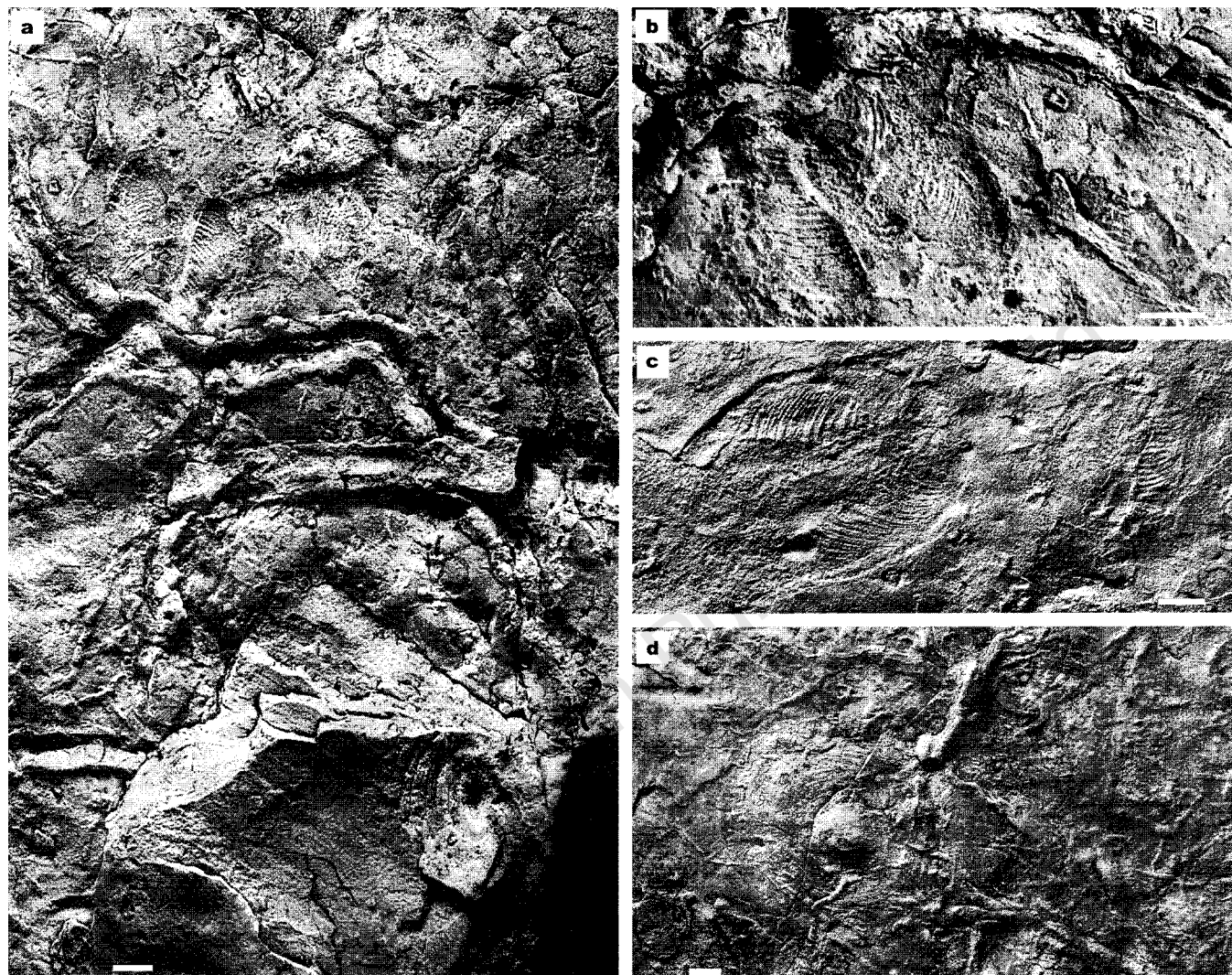


Figure 2 Fronds and associated fossils from the Uratanna Formation, Angepena syncline. Scale bar represents 10 mm. **a**, Upper surface of a fine-grained sandstone-bed with several specimens of *Swartpuntia*-like fronds (new taxon) (South Australian Museum, Adelaide, SAM P36399-36402), and the trace fossil *Taphrhelminthopsis circularis*. An incomplete specimen of the discoidal fossil *Kullingia concentrica* (SAM P36398) is preserved in a bed of fine sandstone

overlying the fronds (lower right). **b**, Best preserved frond (SAM P36399), showing details of lateral elements and nested petaloids. **c**, Two fronds photographed from a latex cast of the upper surface of a fine-grained sandstone bed. **d**, Lower surface of a fine-grained sandstone bed, showing *Kullingia concentrica* (SAM P36394) and the trace fossil *Treptichnus pedum*.

cnidarians and annelids¹ has been challenged on both morphological and taphonomic grounds⁵. The Ediacara-type fossils are preserved in exquisite detail and often with considerable relief, typically on bedding planes of relatively coarse-grained clastic sediments¹⁻⁴. This style of preservation of soft-bodied organisms is effectively unknown in younger sediments. To explain this discrepancy, it was suggested that Ediacaran organisms were members of an extinct kingdom, the Vendobionta, with a unique mattress-like anatomy that is tough enough to fossilize⁵. An alternative view is that Ediacaran preservation reflects unique taphonomic circumstances; this view has led to refinements of the metazoan interpretations¹⁻⁴. The presence of Ediacara-type fossils in Phanerozoic sediments⁷ has been controversial, but a compelling example is the pennatulacean-like *Thaumaptilon* from the Middle Cambrian Burgess Shale⁸. Despite the similarity of *Thaumaptilon* to forms such as *Charniodiscus*, comparisons are not straightforward because of differing styles of preservation⁸.

Here we report the first Phanerozoic occurrence of Ediacara-type fossils showing characteristic coarse-clastic preservation. These fossils are from the Uratanna Formation, from the Angepena

syncline, east of Leigh Creek in the Flinders Ranges, South Australia (Fig. 1). The Uratanna Formation is the basal Lower Cambrian unit in the northern Flinders Ranges, and fills megachannels cut into the underlying terminal Proterozoic Rawnsley Quartzite^{9,10}. The Rawnsley Quartzite contains, in its lower part, the richly fossiliferous Ediacara Member¹¹. Locally, the Uratanna Formation is 375 m thick and forms a gradually, but consistently, shallowing upward sequence from shales to coarse sandstones (Fig. 1).

We found 59 fronds on a surface measuring 40 × 120 cm, and the same bed yielded at least another 20 specimens over a strike length of 400 m. This horizon is about 350 m above the base of the Uratanna Formation, and 25 m below the top of the formation, which is conventionally marked by the appearance of *Diplocraterion*-bearing sandstone of the Parachilna Formation⁹. The newly discovered frondose-body fossils are confined to the top surface of a 10-cm thick, swaley-ripple-topped, fine-grained sandstone, about 3 m below the oldest shoreface facies in the Uratanna Formation. The fronds occur together with prominent horizontal trace fossils *Treptichnus* (formerly *Phycodes*¹²) *pedum* and *Taphrhelminthopsis circularis* (Fig. 2a, b, d). The fossils each consist

of a broad, bilaterally symmetrical frondose body (~4–8 cm long) with a broad central axis (Fig. 2a–c). Sharp arcuate, branchlike elements extend laterally from the axis, curve towards the apex and form narrow to spinose terminations at the margins (Fig. 2c). Each organism seems to have consisted of at least two nested frondose petaloids separated and cast by a thin lens of sand. Although there are many specimens, they do not overlap, indicating that the animals were attached to the sediment and are preserved *in situ*. There is similarity between these fossils and *Swartpuntia* from the terminal Proterozoic Spitskop Member, Namibia¹³, but differences in the tips of the lateral elements and in proportions of axis and vanes show that this is a new taxon.

The Uratanna fronds are known to be Cambrian because they occur with and above an assemblage of trace fossils, including *Treptichnus pedum*, *Taphrhelminthopsis circularis*, *Phycodes* cf. *palmatus*, *Phycodes coronatum* and *Curvolithus* sp.^{9,10}. Trace fossils of the type found in the Uratanna Formation first appear in Cambrian-age sediment¹⁴. In the Newfoundland Precambrian/Cambrian stratotype section at Fortune Head, the boundary is defined by the lowest occurrence of *Treptichnus pedum*¹⁵.

The presence of two specimens of the discoidal fossil *Kullingia concentrica* in the bed immediately overlying the Australian fronds also indicates a Cambrian age for these fronds (Fig. 2a, d). *K. concentrica* occupies a Lower Cambrian level in northern Sweden and the Ukraine^{16,17}. A form of *K. concentrica* also occurs in the Early Cambrian Chapel Island Formation in Newfoundland¹⁸. Organic-walled body fossils (sabeliditids) have been found in the lower part of the Uratanna Formation¹⁰, consistent with, though not diagnostic of, an Early Cambrian age. In all these sections, *K. concentrica* occurs stratigraphically close to sabeliditids and Cambrian-type trace fossils, but below the first unequivocal arthropod trace fossil *Rusophycus* and the vertical spreite burrow *Diplocraterion*^{16–18} (Fig. 1). The unconformity at the base of the Uratanna Formation may reflect a global sea-level change just above or including the Precambrian/Cambrian boundary¹⁹. Other constraints on the precise position of the Precambrian/Cambrian boundary, such as chemostratigraphy and the presence of small shelly fossils, are not available in the Uratanna Formation. Neither, however, are they found in the Newfoundland stratotype section. Nevertheless, the Uratanna Formation can be correlated regionally on the basis of sequence analysis with formations on the Adelaide–Yorke peninsula that contain Cambrian-type small shelly fossils^{20,21}. These results indicate a Nemakit Daldynian age for the Uratanna fronds.

This discovery of *Swartpuntia*-like fronds, within beds burrowed by *T. pedum*, is compelling evidence that Ediacara-type organisms (and preservation of them) extended into the Cambrian period. Until recently it seemed that the Ediacara-type organisms were stratigraphically well separated from the Cambrian stages, and one corollary was the possibility of a major extinction⁶. New discoveries, however, have brought the terminal Proterozoic and Cambrian faunas closer in both time and morphology²². Radiometric data and information from exceptionally complete stratigraphic sections now show that the initial phase of the Cambrian ‘explosion’, corresponding to the Nemakit Daldynian interval, was relatively protracted^{23,24}. Many more late-Proterozoic shelly fossils, including sponges in the form of hexactinellids²⁵ and isolated spicules²⁶, have been discovered. In addition, Ediacara-type fossils are found close to the base of the overlying Cambrian stages in Namibia and Nevada^{13,19}. Whether or not the ‘vendobiont’ hypothesis is true, the existence of the Uratanna fronds shows that there was not a complete destruction of these Ediacara-type forms at the end of the Proterozoic.

A proper understanding of the Ediacara-type fossils will probably depend on taphonomic analysis. Most Ediacara-type organisms lived in shallow-water environments with little sediment bioturbation, and, more significantly, there is indirect evidence that they occurred together with extensive microbial mats^{25,27,28}. Thus, the

unusual terminal Proterozoic preservation could be a product of unique taphonomic conditions involving sediment-binding and -sealing microbial mats, combined with a low degree of bioturbation and scavenging^{25,27,28}. Many of the Ediacaran organisms were benthic forms with limited mobility, or were shallowly placed in the sediment or even attached to the surface of the mat-bound sediment²⁹. The advent of extensive scavenging and sediment disturbance³⁰ would have affected both the biology and the preservation of the Ediacaran organisms. In the early stages of the Cambrian, the extent and depth of bioturbation was still comparatively low³⁰. We suggest that the fossiliferous horizon in the Uratanna Formation offers a glimpse through a rapidly closing taphonomic window. It does not explain the subsequent fate of the different components of the Ediacara fauna, whether they lingered, evolved or perished, but it does demonstrate that Ediacara-type taxa existed in the Cambrian period. □

Received 13 October 1997; accepted 16 March 1998.

- Gehling, J. G. The case for Ediacaran fossil roots to the metazoan tree. *Mem. Geol. Soc. India* **20**, 181–224 (1991).
- Runnegar, B. N. Vendobionta or metazoa? Developments in understanding the Ediacara ‘fauna’. *Neues Jb. Geol. Palaeont. Abh.* **195**, 303–318 (1995).
- Conway Morris, S. The fossil record and the early evolution of the Metazoa. *Nature* **361**, 219–225 (1993).
- Fedonkin, M. A. & Waggoner, B. M. The Late Precambrian fossil *Kimberella* is a mollusc-like bilaterian organism. *Nature* **388**, 868–871 (1997).
- Seilacher, A. Vendobionta and Psammocorallia: lost constructions of Precambrian evolution. *J. Geol. Soc. Lond.* **149**, 607–613 (1992).
- Brasier, M. D. in *Global Events and Event Stratigraphy in the Phanerozoic* (ed. Walliser, O. H.) 113–138 (Springer, Berlin, 1995).
- Crimes, T. P., Insole, A. & Williams, B. J. P. A rigid-bodied Ediacaran Biota from Upper Cambrian Strata in Co. Wexford, Eire. *Geol. J.* **30**, 89–109 (1995).
- Conway Morris, S. Ediacaran-like fossils in Cambrian Burgess Shale-type faunas of North America. *Palaeontology* **36**, 593–635 (1993).
- Daily, B. Discovery and significance of basal Cambrian Uratanna Formation, Mt Scott Range, Flinders Ranges, South Australia. *Search* **4**, 202–205 (1973).
- Mount, J. F. & McDonald, C. Influence of change in climate, sea level, and depositional systems on the fossil record of the Neoproterozoic–early Cambrian metazoan radiation, Australia. *Geology* **20**, 1031–1034 (1992).
- Jenkins, R. J. F., Ford, C. H. & Gehling, J. G. The Ediacara Member of the Rawnley Quartzite: the context of the Ediacara assemblage (Late Precambrian, Flinders Ranges). *Geol. Soc. Aust. J.* **30**, 101–119 (1983).
- Jensen, S. Trace fossils from the Lower Cambrian Mickwitzia sandstone, south-central Sweden. *Fossils Strata* **42**, 1–110 (1997).
- Narbonne, G. M., Saylor, B. Z. & Grotzinger, J. P. The youngest Ediacaran fossils from southern Africa. *J. Paleontol.* **71**, 953–967 (1997).
- Crimes, T. P. Trace fossils and correlation of the late Precambrian and early Cambrian strata. *Geol. Mag.* **124**, 97–119 (1987).
- Narbonne, G. M. & Myrow, P. Trace fossils, small shelly fossils and the Precambrian–Cambrian boundary. *New York State Mus. Bull.* **463**, 72–76 (1988).
- Jensen, S. & Grant, S. W. F. Trace fossils from the Dividal Group, northern Sweden: implications for Early Cambrian biostratigraphy of Baltica. *Norsk Geol. Tids.* (in the press).
- Gureev, Yu. A. in *Problematiki Pozdnego Dokesmbriya i Paleozoya* (eds Sokolov, B. S. & Zhuravleva, I. T.) 92–103 (Nauka, Moscow, 1985).
- Narbonne, G. M., Myrow, P., Landing, E. & Anderson, M. A. A chondrophorine (medusoid hydrozoan) from the basal Cambrian (Placentian) of Newfoundland. *J. Paleontol.* **65**, 186–191 (1991).
- Runnegar, B., Gehling, J. G., Horodyski, R. J., Jensen, S. & Knauth, L. P. Base of the Sauk Sequence is a global eustatic event that lies just above the Precambrian–Cambrian boundary. *Geol. Soc. Am. Abstr.* **27**, 330 (1995).
- Daily, B. Novye dannyye ob osnovanii Kembriya v juzhnoj Avstralii. *Izvestiya Akad. Nauk SSSR, Ser. Geol.* **3**, 45–52 (1976).
- Gravestock, D. I. & Shergold, J. H. in *Ecology of the Cambrian Radiation* (eds Zhuravlev, A. Yu. & Riding, R.) (Columbia Univ. Press, New York, in the press).
- Grotzinger, J. P., Bowring, S. A., Saylor, B. Z. & Kaufman, A. J. Biostratigraphic and geochronologic constraints on early animal evolution. *Science* **270**, 598–604 (1995).
- Bowring, S. A. et al. Calibrating rates of Early Cambrian evolution. *Science* **261**, 1293–1298 (1993).
- Brasier, M. D., Shields, G., Kuleshov, V. N. & Zhegallo, E. A. Integrated chemo- and biostratigraphic calibration of early animal evolution: Neoproterozoic–early Cambrian of southwest Mongolia. *Geol. Mag.* **133**, 445–485 (1996).
- Gehling, J. G. & Rigby, J. K. Long expected sponges from the Neoproterozoic Ediacara Fauna of South Australia. *J. Paleontol.* **70**, 185–195 (1996).
- Brasier, M., Green, O. & Shields, G. Ediacaran sponge spicule cluster from southwestern Mongolia and the origins of the Cambrian fauna. *Geology* **25**, 303–306 (1997).
- Gehling, J. G. A cnidarian of actinian-grade from the Ediacaran Pound Subgroup, South Australia. *Alcheringa* **12**, 299–314 (1988).
- Gehling, J. G. in *12th Int. Sed. Congr. Canberra, Abst.* 117 (1986).
- Seilacher, A. & Pflüger, F. in *Biostabilization of Sediments* (eds Krumben, W. E., Paterson, D. M. & Stal, L. J.) 97–105 (Bibliotheks und Informationssystem Univ. Oldenburg, Oldenburg, Germany, 1994).
- Droser, M. L. & Li, X. in *Ecology of the Cambrian Radiation* (eds Zhuravlev, A. Yu. & Riding, R.) (Columbia Univ. Press, New York, in the press).

Acknowledgements. We thank the National Geographic Society for financial support covering logistics and fieldwork for this project; S. Conway Morris for comments on the manuscript; and M. Walter for reviews. J.G.G. acknowledges financial support from a University of South Australia small ARC grant. S.J. is supported by a grant from NERC.

Correspondence and request of materials should be addressed to S.J. (e-mail: sj10019@esc.cam.ac.uk).

Herbivore-infested plants selectively attract parasitoids

C. M. De Moraes^{*}, W. J. Lewis[†], P. W. Paré[‡], H. T. Alborn[‡] & J. H. Tumlinson[‡]

^{*} Department of Entomology, University of Georgia, Coastal Plain Experiment Station, PO Box 748, Tifton, Georgia 31793, USA

[†] USDA-ARS, IBPMRL, PO Box 748, Tifton, Georgia 31793, USA

[‡] USDA-ARS, CMAVE, PO Box 14565, Gainesville, Florida 32604, USA

In response to insect herbivory, plants synthesize and emit blends of volatile compounds from their damaged and undamaged tissues, which act as important host-location cues for parasitic insects^{1–3}. Here we use chemical and behavioural assays to show that these plant emissions can transmit herbivore-specific information that is detectable by parasitic wasps (parasitoids). Tobacco, cotton and maize plants each produce distinct volatile blends in response to damage by two closely related herbivore species, *Heliothis virescens* and *Helicoverpa zea*. The specialist parasitic wasp *Cardiochiles nigriceps* exploits these differences to distinguish infestation by its host, *H. virescens*, from that by *H. zea*. The production by phylogenetically diverse plant species and the exploitation by parasitoids of highly specific chemical signals, keyed to individual herbivore species, indicates that the interaction between plants and the natural enemies of the herbivores that attack them is more sophisticated than previously realized.

Herbivore-induced plant signals are important for the foraging success of parasitoids and for the plants' defence^{3–8}. The production and release of volatiles is triggered at least in part by substance(s) in the oral secretion of herbivores^{2,9}; in cotton, this is known to be an active process in which several terpenoids are synthesized *de novo* in response to insect feeding¹⁰. The chemical composition of released volatiles varies among plant tissues (cotton leaves, flowers and bolls, for example)¹¹, varieties¹² and cultivars¹³. Although volatile emissions are usually assumed to represent a generalized response, it

has been suggested that plant signals might differ in response to particular herbivore species^{13,14}. To our knowledge, this study represents the first demonstration that herbivore-induced plant volatiles, in the absence of herbivores and damaged tissues, elicit a herbivore-specific natural-enemy response based on quantifiable differences in the proportions of released volatiles.

In field trials (Fig. 1), *C. nigriceps* females distinguished between the odours emanating from tobacco plants infested by their host, *H. virescens*, and those from plants infested by *H. zea* or from undamaged plants. Females selected host over non-host infested plants in 164 of 198 plant visits. This preference persisted when wasps were presented with systematically induced plants from which all damaged leaves were removed (38 of 42 visits over five trials), ensuring that specificity of host preference did not depend on cues associated with the larvae themselves. In all trials involving systematically induced plants, the wasps' response to plants infested by non-hosts was statistically indistinguishable from their response to undamaged control plants. Thus, this parasitoid is able to identify infestations of its host on the basis of chemical cues involving either

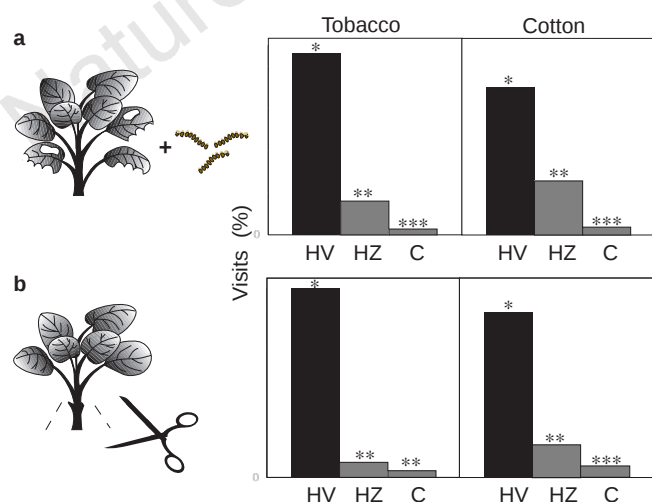


Figure 1 Field-flight responses of *Cardiochiles nigriceps* females to infested tobacco or cotton plants. Plants were infested by larvae of *H. virescens* (HV, tobacco budworm) or *H. zea* (HZ, maize earworm), or were undamaged (C, control), with **a**, caterpillars feeding on the leaves, and **b**, damaged leaves and caterpillars removed from the plant (systematically induced plants)²¹. Asterisks indicate statistically significant differences in the per cent of *C. nigriceps* visits for different herbivore treatments (Duncan's multiple range test after analysis of variance, $n = 5$, $P < 0.05$).

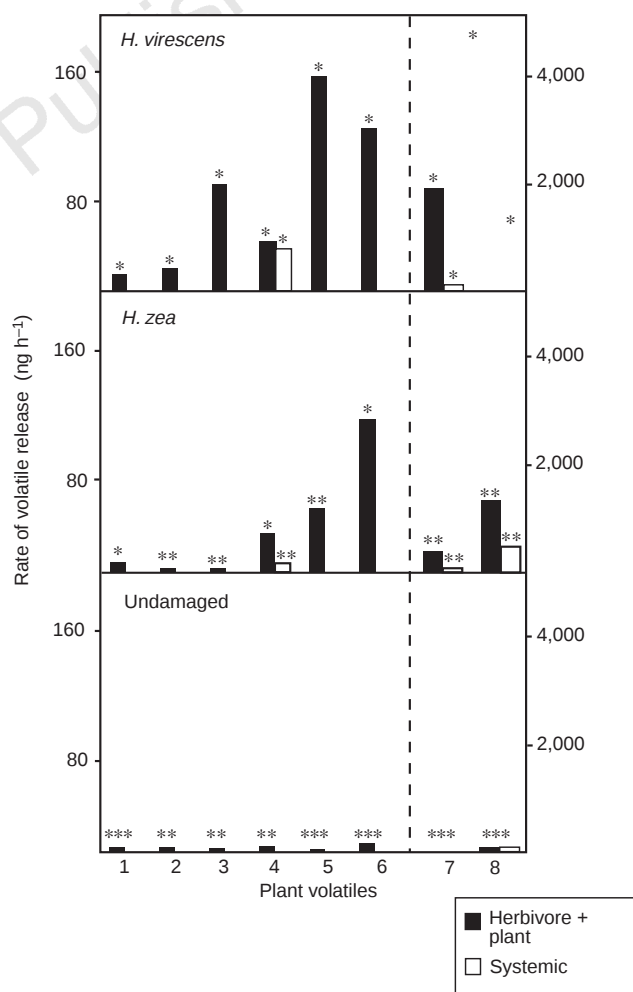


Figure 2 Volatiles released from tobacco plants after 48 h of feeding by *H. virescens* or *H. zea* compared with undamaged controls. Rates of release of volatiles from whole plant/herbivore complexes (black bars) and from systematically induced leaves (white bars) are shown: 1, (Z)-3-hexen-1-ol; 2, linalool; 3, (Z)-Jasmone; 4, α -humulene; 5, (*E,E*)- α -farnesene; 6, unidentified sesquiterpene; note the different scale for compound 7, (*E*)- β -ocimene, and compound 8, β -caryophyllene, present in much higher quantities than other components. Asterisks indicate statistically significant differences in the average amount released from 12:00 to 15:00 h for different treatments (Duncan's multiple range test after analyses of variance, $n = 4$, $P < 0.05$).

quantitative or qualitative differences in the composition of volatile compounds produced systematically and emitted by the plant.

Gas chromatography revealed consistent differences between the proportions of volatile compounds released by tobacco plants damaged by *H. virescens* and those damaged by *H. zea*. Eight major compounds were consistently released by both of these herbivore-plant complexes. Of these, concentrations of five were significantly higher in complexes involving *H. virescens* than in those involving *H. zea* (Fig. 2), indicating a difference in the relative proportions of compounds in each blend. There was no significant difference in the overall average amount of damage caused by *H. virescens* ($3,842 \pm 548 \text{ mm}^2$ per plant) and *H. zea* ($3,878 \pm 552 \text{ mm}^2$ per plant); in some cases herbivore numbers were manipulated to ensure consistent feeding damage (some differences in rates of damage within the feeding period cannot be fully discarded). When systematically induced volatiles alone were examined, concentrations of three compounds were significantly higher in *H. virescens*- than in *H. zea*-damaged plants (Fig. 2). In particular, 48 h after initial herbivore damage, the systemic release of β -caryophyllene and β -ocimene from tobacco plants damaged by *H. virescens* was raised three- and five-fold, respectively, over that from plants damaged by *H. zea*. Thus, the qualitative composition of volatile blends emitted by tobacco plants in response to specific herbivores can differ significantly and consistently, providing a reliable signal by which *C. nigriceps* may identify *H. virescens*-infested plants.

C. nigriceps preference for *H. virescens*-damaged plants might still be explained by increased volatile emissions from *H. virescens*-damaged tobacco plants rather than by qualitative differences between volatile blends. We therefore repeated the behavioural and chemical studies using cotton plants (*Gossypium hirsutum*) damaged by *H. virescens* and *H. zea* (in separate treatments) to determine whether host-specific responses by *C. nigriceps* depend on herbivore-specific signalling or merely on raised volatile emis-

sions associated with *H. virescens*-damaged tobacco. Indigenous populations of both herbivore species attack cotton and occur together with *C. nigriceps* on a regular basis, so these tests approximated a natural situation. Moreover, similar amounts of volatile emissions by cotton in response to both herbivore species were anticipated. Thus, this experiment provided a control for a potentially important source of variation and confirmed that the low level of volatile emissions from damaged tobacco plants was not due merely to the absence or attenuation of a plant-volatile elicitor in oral secretions of *H. zea*. In field choice bioassays, *C. nigriceps* demonstrated a significant preference for cotton plants damaged by *H. virescens* over those damaged by *H. zea* (49 of 70 visits over 3 trials). This preference persisted when wasps were presented with systematically induced plants from which all damaged leaves and insects were removed (19 of 25 visits over 3 trials) (Fig. 1). Thus, *C. nigriceps* can distinguish cotton as well as tobacco plants damaged by *H. virescens* from those damaged by *H. zea* solely on the basis of information from undamaged portions of damaged plants.

Gas chromatography of volatile compounds released by herbivore-cotton complexes indicated that there were consistent differences in plant responses to the two herbivores (Fig. 3). No significant difference was observed in the total volatile amount released by the two plant-herbivore complexes. However, significant and consistent differences were found in the qualitative composition of the volatile blends emitted by cotton plants in response to similar feeding by *H. virescens* as compared to *H. zea*.

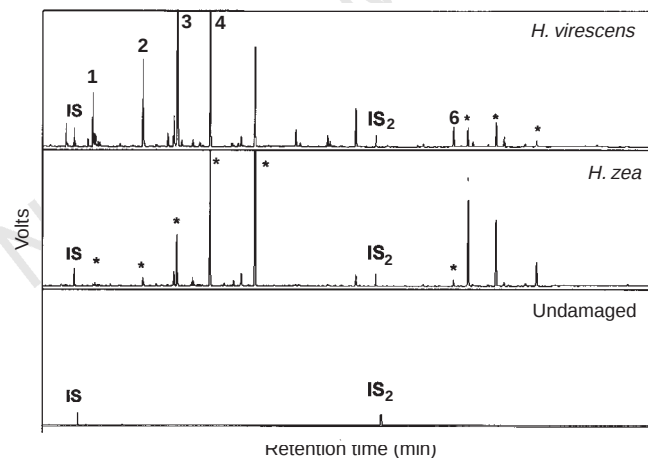


Figure 3 Chromatographic profiles of volatiles from cotton after 24 h of feeding by *H. virescens* or *H. zea* compared with undamaged controls. Significant differences are indicated by numbers, with identities and emission levels (ng per 3 h) as follows (*H. virescens*/*H. zea*): 1, (Z)-3-hexen-1-ol (13,738/215); 2, α -pinene (63,310/748.0); 3, (Z)-3-hexenyl acetate (111,350/6,922); 4, (E)- β -ocimene (120,257/5,919); 5, (E)-4,8-dimethyl-1,3,7-nonatriene (35,612/88,000); 6, β -caryophyllene (5,591/546); 7, (E)- β -farnesene (HV: 9,177/247); 8, (E,E)- α -farnesene (1,272.6/9,072.6); 9, (E,E)-4,8,12-trimethyl-1,3,7,11-tridecatetraene (3,075/9,317). Internal standards (IS, *n*-octane; IS₂, nonyl acetate). Considerable variation over the 48-h collection period included occurrence of other compounds, but differences in reported compounds were consistent throughout (Duncan's multiple range test after analyses of variance, $n = 3$, $P < 0.05$). Asterisks designate compounds that align with numbered peaks in the corresponding chromatogram.

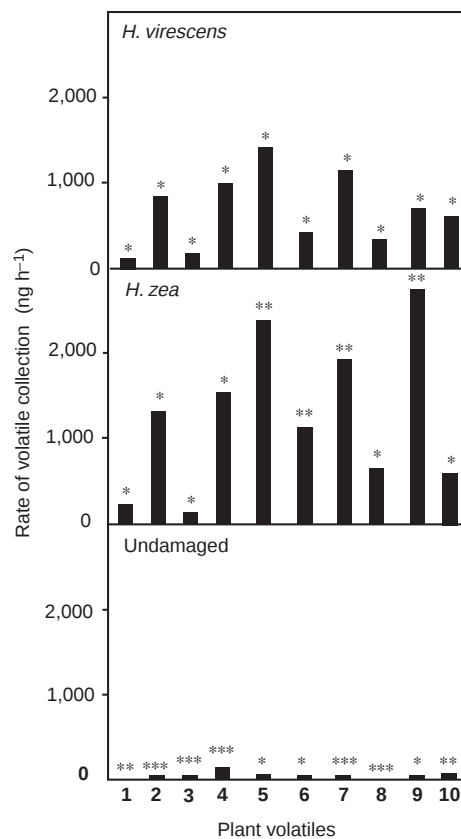


Figure 4 Volatiles collected from maize seedlings (*Zea mays*, strain LG11) fed on for 24 h by *H. virescens* or *H. zea* or from undamaged control plants. Bars in each graph represent the following compounds: 1, (Z)-3-hexen-1-ol; 2, (Z)-3-hexenyl acetate; 3, (E)- β -ocimene; 4, linalool; 5, (E)-4,8-dimethyl-1,3,7-nonatriene; 6, indole; 7, β -caryophyllene; 8, α -trans bergamotene; 9, (E)- β -farnesene; 10, (E,E)-4,8,12-trimethyl-1,3,7,11-tridecatetraene. Asterisks indicate statistically significant differences in the average amount released from 12:00 to 15:00 h for different treatments (Duncan's multiple range test after analyses of variance, $n = 5$, $P < 0.05$).

Concentrations of five compounds were significantly higher in samples from *H. virescens*-damaged plants. On the other hand, concentrations of four compounds were raised in the *H. zea* complex. As seven of these compounds (two from *H. virescens* and four from *H. zea*) are biosynthesized *de novo* in response to insect feeding on cotton leaves⁹, plants must produce qualitatively distinct volatile blends in response to feeding by different herbivore species. Moreover, herbivore-specific volatile blends differ significantly and consistently, providing reliable, information-rich signals to appropriate natural enemies. These results, together with behavioural data (Fig. 1), confirm that the parasitoid *C. nigriceps* uses such plant signals to distinguish infestations of its host, *H. virescens*, from those of the closely related *H. zea*.

Repetition of our analysis using maize provided further evidence that herbivore-specific signalling is common to a wide variety of plants. Chromatography of maize-released volatiles revealed consistent differences in plant responses to the two herbivores (Fig. 4): *H. zea* elicited a significant increase in the release (compared to *H. virescens*) of four of the ten main compounds that were consistently present from damaged maize (overall average damage by *H. virescens*, $2,542 \pm 378 \text{ mm}^2$, and by *H. zea*, $2,601 \pm 401 \text{ mm}^2$). Our observation that herbivore-specific plant volatile responses occur among several plant families (Malvaceae, Solanaceae and Gramineae) indicates that this phenomenon is widely spread in the plant kingdom. The proportions of released compounds, and in some cases the compounds themselves, vary across plants.

By exploiting herbivore-specific volatile emissions, *C. nigriceps* is able to distinguish between host and non-host infestations on phylogenetically distant plant species that produce varying chemical blends (that is, volatile signals vary across plant species for each herbivore). This ability to distinguish host from non-host using different signals on different plants implies a level of behavioural plasticity on the part of these parasitoids, in keeping with demonstrations that parasitoid behavioural repertoires are flexible, involving learning and context-dependent analysis of chemical cues^{11,15,16}. This perspective contrasts with the classical view of rigid behavioural responses mediated by pheromonal systems and tightly linked to insect physiology through a 'hard-wired' arrangement¹⁷. As indicated by herbivore-specific plant signalling, parasitoids can exploit diverse chemical cues in a context-dependent manner on the basis of information content. Such a keen-yet-flexible detection system allows a specialist parasitoid like *C. nigriceps* to locate its polyphagous host over a range of potential habitats.

Our demonstration that parasitoids can discriminate, on the basis of systematically produced plant volatile profiles, between plants that have been fed upon by their hosts and those that have been attacked by non-hosts is evidence of a sophisticated system of herbivore-specific plant/parasitoid chemical interactions. By providing specific and reliable chemical signals, plants may obtain a competitive advantage in the acquisition of herbivore natural enemies^{18,19,20}. Parasitoids may concurrently improve their foraging efficiency by using these signals to distinguish host from non-host infestation. We now know that tobacco, maize and cotton plants produce herbivore-specific chemical signals in response to herbivory and that the specialist parasitic wasp *C. nigriceps* can exploit such information-rich signals in locating hosts. Moreover, these wasps exhibit a flexible behavioural response to different signals produced by phylogenetically distant plant families. These observations advance our understanding of the complex tritrophic interactions among plants, herbivores and parasitoids. □

Methods

Behavioural assays. Twelve laboratory-reared, third-instar caterpillars were caged on the lower leaves (two caterpillars per leaf) of 8-week-old, potted, greenhouse-grown tobacco (*Nicotiana tabacum* strain K326) or cotton (*Gossypium hirsutum* strain DPL 90) plants for 48 hours before placement in

the field. In separate experiments, six cotton or six tobacco plants (two damaged by *H. virescens*, two by *H. zea*, and two undamaged) were placed in a cotton field with an active population of *C. nigriceps*. The plants were arranged 80 cm apart in a two-by-two design, with the two treatments placed in alternate positions. The number of female parasitoids landing on each plant in a period of one hour was visually observed and recorded. Each bioassay was conducted on five days to account for day-to-day variation. For assays measuring effects of systematically induced plants²¹, the damaged leaves were removed and the cut areas covered with aluminium foil. Preference between the plants damaged by the two herbivores and undamaged plants was subjected to ANOVA and means were compared by an LSD test using a statistical analysis system (SAS Institute, Cary, NC).

Volatile compound collection. Ten third-instar larvae of either *H. zea* or *H. virescens* were allowed to feed on the top four leaves of 8-week-old, greenhouse-grown tobacco plants, the top five leaves of 6-week-old, greenhouse-grown cotton plants, or the leaves of three 10-day-old maize seedlings (*Zea mays* strain LG11) enclosed in glass sleeves. Volatile collection started 48 h after initiation of feeding on the tobacco plants and 24 h after initiation of feeding on the cotton and maize plants. Caterpillars were allowed to continue feeding during volatile collection.

For systemic treatments, two caterpillar per leaf for a total of five leaves were enclosed on the lower part of the plant in cages consisting of half Petri dishes. In control experiments, empty cages were used. The upper section of each plant (without feeding damage) was placed in the glass sleeve ($34 \times 14 \text{ cm}$ diameter for tobacco and cotton plants and $34 \times 7 \text{ cm}$ diameter for maize plants). A split plate at the base of the sleeve closed loosely around the stem of the plant¹⁹, allowing the infested leaves to remain outside the chamber to avoid contamination²².

Filtered air (1 litre min^{-1}) was passed from the top down over the plants and out of a hole in the plate. Plant volatiles were collected between 15:00 (day 1) and 18:00 (day 2) at 3-h intervals by pulling one fifth (tobacco and cotton plants) and one half (maize plants) of the air passing over the plant through Super-Q adsorbent (25 mg) traps at the base of the volatile-collection chamber; the remainder of the air vented out of the bottom of the system. The traps were rinsed with 150 μl methylene chloride and 400 ng of *n*-octane and nonyl acetate were added as internal standards. 1- μl aliquots of the samples were injected, using a splitless injector held at 240 °C, into a Hewlett–Packard model 5890 gas chromatograph. The OV 101 methyl silicon column (Quadrex) was held at 60 °C for 2 min and then programmed at 4 °C per min to 180 °C, where it was maintained for 10 min. A PE Nelson data collection system and Turbochrom 3.1 software (Perkin Elmer) was used to quantify all major components based on the response to the internal standards. The structure of the volatile components was confirmed using commercially available standards and GC/MS analysis.

Leaf measurement. The leaves on which larvae had been feeding were removed from the plants and photocopied. These photocopies were scanned and imported into an imaging program (ImagePC beta version, Scion) by which leaf area eaten and leaf remaining were quantified.

Received 21 November 1997; accepted 16 March 1998.

1. Dicke, M. Local and systemic production of volatile herbivore-induced terpenoids: their role in plant-carnivore mutualism. *J. Plant Physiol.* **143**, 465–472 (1994).
2. Turlings, T. C. J., Tumlinson, J. H. & Lewis, W. J. Exploitation of herbivore-induced plant odors by host-seeking parasitic wasps. *Science* **250**, 1251–1253 (1990).
3. McCall, P. J., Turlings, T. C. J., Lewis, W. J. & Tumlinson, J. H. Role of plant volatiles in host location by the specialist parasitoid *Microplitis croceipes* Cresson (Braconidae:Hymenoptera). *J. Ins. Behav.* **6**, 625–639 (1993).
4. Karban, R. Resistance against spider mites in cotton induced by mechanical abrasion. *Entomol. Exp. appl.* **37**, 137–141 (1985).
5. Dicke, M. *et al.* Isolation and identification of volatile kairomone that affects acarine predator–prey interactions. Involvement of host plant in its production. *J. Chem. Ecol.* **16**, 381–396 (1990).
6. Mattiacci, L., Dicke, M. & Posthumus, M. A. Induction of parasitoid attracting synomone in brussels sprouts plants by feeding of *Pieris brassicae* larvae: Role of mechanical damage and herbivore elicitor. *J. Chem. Ecol.* **20**, 2229–2247 (1994).
7. Turlings, T. C. J. *et al.* How caterpillar-damaged plants protect themselves by attracting parasitic wasps. *Proc. Natl Acad. Sci. USA* **92**, 4169–4174 (1995).
8. Alborn, H. T. *et al.* An elicitor of plant volatiles from beet armyworm oral secretion. *Science* **276**, 945–948 (1997).
9. Paré, P. W. & Tumlinson, J. H. Induced synthesis of plant volatiles. *Nature* **385**, 30–31 (1997).
10. Udayagiri, S. & Jones, R. L. Role of plant odor in parasitism of European corn borer by braconid specialist *Macrocentrus grandii* Goidanich: Isolation and characterization of plant synomones eliciting parasitoid flight response. *J. Chem. Ecol.* **18**, 1841–1855 (1992).
11. Turlings, T. C. J., Wäckers, F. L., Vet, L. E. M., Lewis, W. J. & Tumlinson, J. H. In *Insect Learning* (eds Papaj, D. R. & Lewis, A. C.) 51–78 (Chapman & Hall, New York, 1993).

12. Loughrin, J. H., Manukian, A., Heath, R. R. & Tumlinson, J. H. Diurnal cycle of emission of induced volatile terpenoids by herbivore-injured cotton plants. *J. Chem. Ecol.* **21**, 1217–1227 (1994).
13. Takabayashi, J., Dicke, M. & Posthumus, M. A. Variation in composition of predator-attracting allelochemicals emitted by herbivore-infested plants: Relative influence of plant and herbivore. *J. Chem. Ecol.* **22**, 1591–1605 (1996).
14. Du, Y.-J., Poppy, G. M. & Powell, W. Relative importance of semiochemicals from first and second trophic levels in host foraging behavior of *Aphidius ervi*. *J. Chem. Ecol.* **22**, 1591–1605 (1996).
15. Lewis, W. J. & Takasu, K. Use of learned odours by a parasitic wasp in accordance with host and food needs. *Nature* **348**, 635–636 (1990).
16. Tumlinson, J. H., Lewis, W. J. & Vet, L. E. M. How parasitic wasps find their hosts. *Sci. Am.* **268**, 100–106 (1993).
17. Bell, W. J., Kipp, L. R. & Collins, R. D. in *Chemical Ecology of Insects 2* (eds Cardé, R. T. & Bell, W. J.) 105–154 (Chapman & Hall, New York, 1995).
18. Strand, M. R. & Obrycki, J. J. Host specificity of insect parasitoids and predators. *BioScience* **46**, 422–429 (1996).
19. Futuyma, D. J. & Moreno, G. The evolution of ecological specialization. *Annu. Rev. Ecol. Syst.* **19**, 207–233 (1988).
20. Thompson, J. N. *The Coevolutionary Process* (Univ. of Chicago Press, Chicago, 1994).
21. Rose, U. S. R., Manukian, A., Heath, R. R. & Tumlinson, J. H. Volatile semiochemicals released from undamaged cotton leaves: A systemic response of living plants to caterpillar damage. *Plant Physiol.* **111**, 487–495 (1996).
22. Heath, R. R. & Manukian, A. *J. Chem. Ecol.* **20**, 593–608 (1994).

Acknowledgements. We thank M. C. Mescher for assistance in preparing the manuscript; T. C. Turlings, P. J. Landolt, J. Garcia, B. Benrey, K. Ross, K. Korth and J. Ruberson for comments on the manuscript; A. T. Proveaux for assistance with mass spectrometric analysis; J. H. Loughrin for leaf measurement software; and T. Green for insect rearing. C.D.M. is recipient of fellowship from CAPES (Brazil).

Correspondence and requests for materials should be addressed to W.J.L. (e-mail: WJL@tifton.cpes.peachnet.edu).

Evolution of indirect reciprocity by image scoring

Martin A. Nowak^{*†} & Karl Sigmund[‡]

^{*} Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

[‡] Institut für Mathematik, Universität Wien, Strudlhofgasse 4, A-1090 Wien, Austria

Darwinian evolution has to provide an explanation for cooperative behaviour. Theories of cooperation are based on kin selection (dependent on genetic relatedness)^{1,2}, group selection^{3–5} and reciprocal altruism^{6–9}. The idea of reciprocal altruism usually involves direct reciprocity: repeated encounters between the same individuals allow for the return of an altruistic act by the recipient^{10–16}. Here we present a new theoretical framework, which is based on indirect reciprocity¹⁷ and does not require the same two individuals ever to meet again. Individual selection can nevertheless favour cooperative strategies directed towards recipients that have helped others in the past. Cooperation pays because it confers the image of a valuable community member to the cooperating individual. We present computer simulations and analytic models that specify the conditions required for evolutionary stability¹⁸ of indirect reciprocity. We show that the probability of knowing the ‘image’ of the recipient must exceed the cost-to-benefit ratio of the altruistic act. We propose that the emergence of indirect reciprocity was a decisive step for the evolution of human societies.

Humans have achieved one of the pinnacles of sociality, and the complexity of their cooperative actions is without parallel. In contrast to other examples of ultrasociality^{19–22} (for example, clones, bee hives or termite colonies), human cooperation is due less to kin selection than to cultural forces rooted in pervasive moral systems. From hunter tribes and village communities to nation states and global enterprises, the economic effects of nepotism, although certainly present, are minor compared with those of reciprocity. Reciprocity is usually understood to take the form of direct reciprocity: help someone who may later help you. But indirect reciprocity also prevails in human communities. In this case, one does not expect a return from the recipient, but from someone else, according to the pious advice of ‘give, and you shall be

given’. Cooperation is channelled towards the ‘valuable’ members of the community. This has been called the ‘I won’t scratch your back if you won’t scratch their backs’ principle²³. A donor provides help if the recipient is likely to help others (which often means, if the recipient has helped others in the past). In this case, it pays to advertise cooperation, as the cost of an altruistic act is offset by an increased chance to become the recipient of an altruistic act later. Animal and human behaviour may be influenced by attempting to increase image (or status) in the group^{24,25}.

According to Alexander¹⁷, indirect reciprocity, which ‘involves reputation and status, and results in everyone in the group continually being assessed and reassessed’, is important in human societies (and possibly in some primates, social canines and other groups). Alexander interprets moral systems as systems of indirect reciprocity. Indirect reciprocity presupposes rather sophisticated players, and therefore is likely to be affected by anticipation, planning, deception and manipulation. The politicking needed to

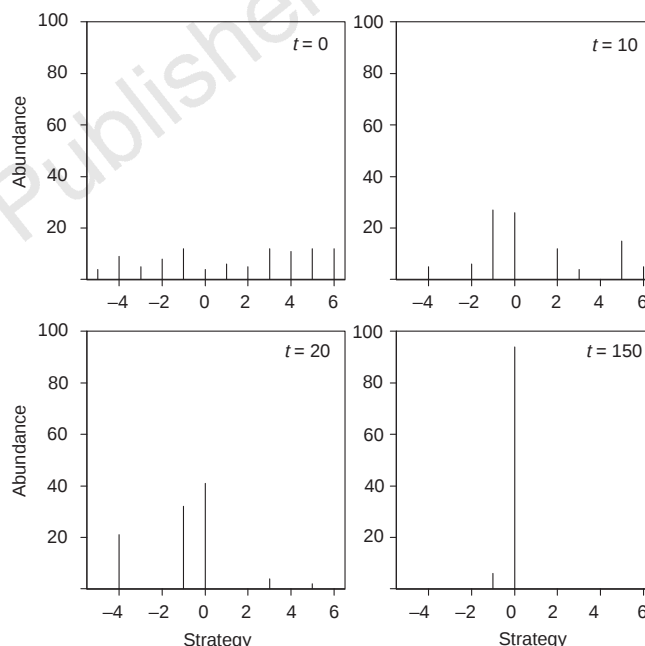


Figure 1 Cooperation wins in a computer simulation of indirect reciprocity. The population consists of $n = 100$ individuals. The image scores range from -5 to $+5$, the strategy (k) values from -5 to $+6$. The strategy $k = -5$ represents unconditional cooperators, whereas the strategy $k = +6$ represents defectors. In each round of the game, two individuals are chosen at random, one as donor, the other as recipient. The donor cooperates if the image score of the recipient is greater than or equal to the donor's k value. Cooperation means the donor pays a cost, c , and the recipient obtains a benefit, b . There is no payoff in the absence of cooperation. At the beginning of each generation, all players have image score 0. Hence, strategies with $k \leq 0$ are termed ‘cooperative’, because individuals with these strategies cooperate with individuals that have not had an interaction. In each generation 125 donor–recipient pairs (n) are chosen; each player has, on average, 2.5 interactions. The chance that a given player meets the same player again, or that a chain of possible altruistic acts ever leads back to the original donor, is negligibly small. Therefore, direct reciprocity cannot work here. At the end of each generation, players produce offspring proportional to their payoff. At generation, $t = 0$, we start with a random distribution of strategies. After $t = 10$ generations, the strategies $k = -1, 0, +2$ and $+5$ have increased in abundance. After $t = 20$ generations, the strategies $k = -4, -1$ and 0 dominate the population. After $t = 150$ generations, the population consists almost entirely of the strategy $k = 0$, which is the most discriminating among all cooperative strategies. Players with this strategy cooperate with everyone who has image score 0 or greater. After $t = 166$ generations, all other strategies have become extinct and $k = 0$ is fixed in the population. Parameter values: $b = 1, c = 0.1$ (to avoid negative payoffs we add 0.1 in each interaction).

[†] Present address: Institute for Advanced Study, Olden Lane, Princeton, New Jersey 08540, USA.

continually assess the status of all members of our community and to bolster our own has probably been a major force for shaping our intelligence. But if we want to understand the basic mechanisms of indirect reciprocity, we must analyse drastically simplified models.

Imagine a population of individuals with the option to help one another or not. Random pairs of players are chosen, of which one is the potential donor of some altruistic act and the other is the recipient. The donor can cooperate and help the recipient at a cost c to himself, in which case the recipient receives a benefit of value b (with $b > c$). If the donor decides not to help, both individuals receive zero pay-off. Each player has an image score, s , which is known to every other player. If a player is chosen as a donor and decides to cooperate then his (or her) image score increases by one unit; if the donor does not cooperate then it decreases by one unit. The image score of a recipient does not change. First we consider strategies where donors decide to help according to the image score of the recipient. A strategy is given by a number k : a player with this

strategy provides help if, and only if, the image score of the potential recipient is at least k .

Figure 1 shows computer simulations of a population consisting of n players. The strategies are given by k_i and the image levels by s_i . At the beginning of each generation, the image levels of all players are zero (assuming that children do not inherit the image of their parents). In succession, m donor-recipient pairs are chosen. A donor, i , cooperates with a recipient, j , if $k_i \leq s_j$. The fitness of a player is given by the total number of points received during the m interactions. Some players may never be chosen, in which case their payoff from the game will be zero. On average, a player will be chosen $2m/n$ times, either as donor or as recipient. At the end of each generation, players leave offspring in proportion to their fitness. We find that if the game is played for many generations, then eventually all players will adopt the same strategy. If the k value of this strategy is 0 or less then cooperation is established; if the value is 1 or more then defection has won. Cooperation is more likely to win if the number of interactions, m , per generation is large. (A different model of indirect reciprocity has been studied by Boyd and Richerson²⁶, who assumed that individuals interact in loops such that a cooperative action can be returned, after several steps, to the original donor. According to Boyd and Richerson, their model is unlikely to lead to a cooperative outcome, as it requires the loops to

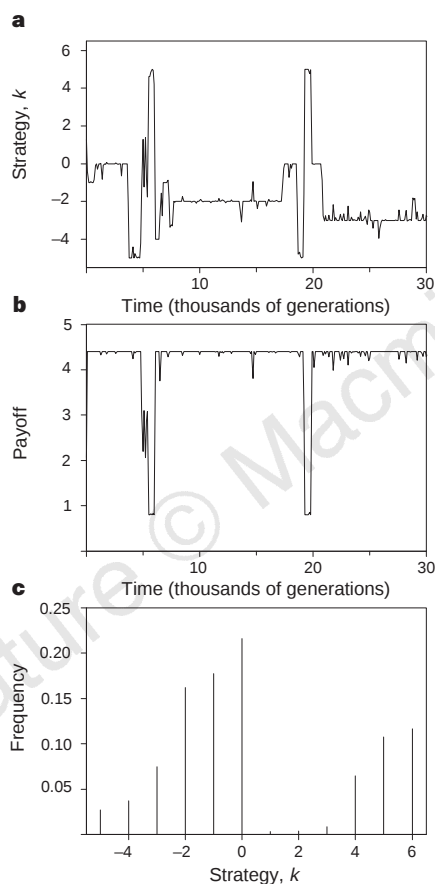


Figure 2 Long-term evolution of indirect reciprocity under mutation and selection. We used the same computer simulation as in Fig. 1, but included mutation: there is a probability of 0.001 that an offspring does not act like its parent and uses another randomly chosen strategy instead. We observe endless cycles of cooperation and defection. Cooperative populations are relatively stable if they consist of discriminating players with strategies such as $k = 0$ or -1 . But after some time these populations are undermined (through random drift) by players with strategies such as $k = -4$ or -5 , which are too cooperative. Then defectors, with strategies $k = 4$ or 5 , can invade. These defectors can, in turn, be overcome by stern discriminators again. In the long run, cooperation is harmed by unconditional cooperators, because they enable defectors to invade. In the absence of unconditional cooperators, cooperative populations persist for much longer. **a**, The average k value of the population. **b**, The average payoff per individual, per generation. **c**, Frequency distribution of strategies sampled over many generations ($t = 10^7$). Parameter values are as for Fig. 1, but $m = 300$ rounds per generation.

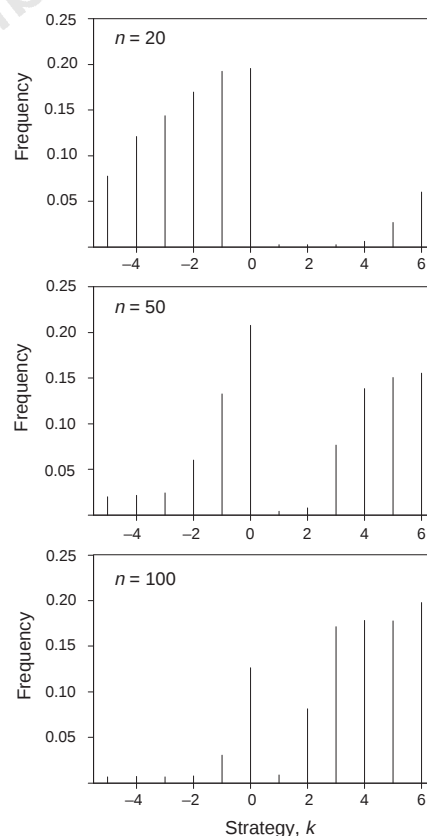


Figure 3 Indirect reciprocity with incomplete information about the image score of other players. We performed the same simulation as in Fig. 2, but updated the image score of a donor only for the recipient and for the observers of an interaction. Each interaction is observed, on average, by ten randomly chosen players. The figure shows the frequency distribution of strategies for three different population sizes, $n = 20$, $n = 50$ and $n = 100$, sampled over many generations ($t = 10^7$) in order to obtain representative results. There is a clear effect of group size: cooperation predominates for $n = 20$, but is rare for $n = 100$. For $n = 50$ we find cooperative and defective strategies at roughly equal frequencies. The averages over time of the frequency of cooperative strategies (defined by $k \leq 0$) are 90%, 47% and 18% for, respectively, $n = 20$, 50 and 100. Parameter values are as for Fig. 2, but the number of donor-recipient interactions, m , is $10n$ rounds per generation.

be relatively small, closed and long lasting. We think that this is because their model does not include image scores.)

We can also include mutations in the simulation, by assuming that there is a small probability that a strategy does not reproduce accurately but instead gives rise to an offspring adopting a different strategy (Fig. 2). In this case, several strategies can persist. We have studied the frequency distribution of various strategies and analysed how often a cooperative regime is achieved. A minimum number of rounds per generation is needed for cooperation to prevail. This number can be very small: each player needs to be chosen for only about two interactions per lifetime. (In this case, there is a probability of only 1/4 that a defector can be punished, if he is chosen first as a donor and then as a recipient.) Below we present an

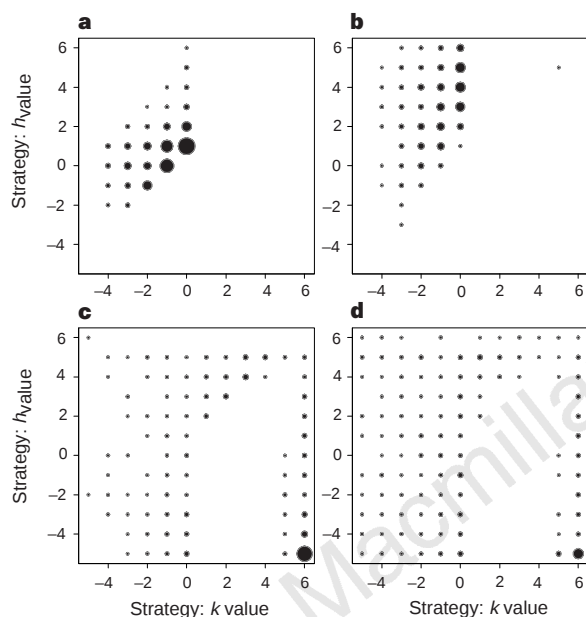


Figure 4 A further dimension is added to the game if donors base their decision to cooperate not only on the image score of the recipient but also on their own score.

a, b, We consider strategists that cooperate if the image score of the opponent is at least k and if their own image score is less than h . If the image score of an individual is already high, it makes no sense to invest in a still higher image. The figures show the frequency distribution of strategies that are defined by their k and h values sampled over many generations. **a,** We assume perfect information about the image of all players. The most frequent strategy is $k = 0, h = 1$. Players with this strategy cooperate if the image score of the opponent is at least 0 and their own image score is less than 1. If the whole population adopts this strategy, it does not pay to aim for an image exceeding 0. For the same reason, other strategies with $h = k + 1$ are successful in this simulation. Strategies with $k > 0$ are unsuccessful, because they are too uncooperative. **b,** We assume imperfect information about the other players' image. Here it pays to invest in a higher image than strictly necessary, because a given altruistic act is only seen by a subset of other players. The most frequent strategy is $k = 0, h = 4$. **c, d,** We study players that cooperate if the image score of the recipient is at least k or if their own image score is less than h . Players with a low image may want to increase their image by helping others indiscriminately. Such a scenario also leads to cooperative societies (dominated by strategies with $k \leq 0$), but unconditional defectors (with strategies $k = 6, h = -5$) benefit from the reduced level of discrimination and represent the most frequent single strategy. **c,** Results are based on perfect information whereas **d** assumes imperfect information about the co-players' image score. In **a, b, c** and **d**, respectively, the frequency of cooperative interactions is 55%, 57%, 70% and 80%. This must be compared with $<0.1\%$ cooperation in simulations where strategies only consider their own image score and do not discriminate according to the image score of the recipient. Parameter values: **a, c,** as in Fig. 2, but $m = 500$ rounds per generation; **b, d,** as in Fig. 3 with $n = 20$. The frequency of a strategy is proportional to the area of the shaded circles. Strategies with a frequency of $<0.5\%$ are not shown.

analytical model for evaluating the minimum number of interactions that is compatible with cooperation.

Long-term simulations that include mutation usually do not converge to a simple equilibrium distribution of strategies, but show endless cycles. In simple terms, what happens is that defectors are invaded by discriminators, who only help players whose score exceeds some threshold. Next, discriminators are undermined by unconditional cooperators. The prevalence of these indiscriminate altruists subsequently allows the return of defectors. In a population consisting only of discriminators and unconditional cooperators, there is no selection against the latter, who can spread by random drift. In simulations without unconditional cooperators, cooperative populations persist for much longer.

Cooperation based on indirect reciprocity depends crucially on the ability of a player to estimate the image score of the opponent. In the above model, we assume that the image score of each individual is known to every other member of the population. This should be seen as only an idealized scenario. It is more realistic to assume that an interaction between two individuals is observed by a (possibly small) subset of the population. Only these 'onlookers' (and, of course, the recipient) have the possibility of updating their perception of the donor's image score. The onlookers are chosen at random for each particular interaction. Therefore each player has a specific perception of the image score of the other players. The same player can have different image scores in the eyes of different individuals. The information is contained in a matrix whose elements s_{ij} denote the image score of player i as seen by player j . In a donor–recipient interaction between j and i , player j will cooperate if $s_{ij} > k_j$. If j has no information on i , then $s_{ij} = 0$.

The model now depends on the probability that a given individual observes an interaction between two other individuals. Figure 3 shows computer simulations of this extended model. Again, cooperation can easily be established and dominate the population, but a larger number of interactions per generation is needed. There is also an effect of group size. For larger groups, it is more difficult to establish cooperation, because the fraction of individuals that obtain information about any particular interaction will be smaller. Therefore, more interactions are required (relative to group size) in order to discriminate against defectors.

Another interesting expansion of the basic model is to include strategies that consider both the recipient's and the donor's image score. We explored two types of strategies. 'And' strategies involve cooperation if the image score of the recipient is larger than a certain value and the image score of the donor is less than a certain value. The idea is that if an individual has already a high image score, it is not necessary to aim for a still higher image score (by helping others). On the other hand, 'or' strategies result in cooperation if the image score of the recipient is larger than a certain value or the image score of the donor is less than a certain value. Here the idea is that if an individual has a low image score it may be advantageous to increase the score by helping others regardless of how low their image score is. In both cases, highly cooperative societies form (Fig 4). If, in contrast, we simulate strategies that only consider their own image and do not take into account the image of the recipient, cooperation does not emerge.

The models above are based on computer simulations, but we can derive analytical insights from a simplified model. Suppose that there are only two image levels, 0 (for bad) and 1 (for good). The image of a player depends on his or her last action as a donor: players who defected have score 0, and players who cooperated have score 1. Let us only consider two types of player: first, defectors, who never provide assistance; and second, discriminators who help players having image 1, but not players having image 0. A given player knows the score of only a fraction, q , of the population. A discriminator who has no information on potential recipients will assume, with a certain probability, p , that they have image 1. In each round of the game all individuals of the population are chosen, each

with the same probability of being a donor or a recipient. If $w < 1$ denotes the probability of another round, there are on average $1/(1-w)$ rounds per generation. We have derived the equations (see Methods) that describe how the frequencies of discriminators and defectors change from one generation to the next. It should be stressed that discriminators are not 'tit-for-tat' players; tit-for-tat strategists base their decisions on their own previous experience with the co-player, whereas discriminators use the experience of others. This is an essential advantage for a player who interacts with many co-players but only a few times with each. (Such discriminators are also different from strategies based on 'standing'²⁷, which is an internal switch distinguishing between defection in response to a co-player's cooperation or defection.)

We observe a frequency threshold: a minimum amount, $x_{\min}$, of discriminators is necessary to ensure the establishment of cooperation. We also obtain the minimum number of rounds per generation that is needed for the evolutionary stability of discriminators. In particular, cooperation through indirect reciprocity can only be stable if $q > c/b$. The probability of knowing the image of another player has to exceed the cost-to-benefit ratio of the altruistic act. This is remarkably similar to Hamilton's rule, which states that cooperation through kin selection works whenever the coefficient of relatedness between two individuals exceeds the cost-to-benefit ratio^{1,2}. In our case, relatedness is replaced by acquaintanceship.

Cooperation based on indirect reciprocity works in the following way, therefore: a potential donor can choose whether to accept a certain cost in order to help another individual, or to avoid this cost. In the short term, of course, avoiding the cost yields the higher payoff. In the long term, however, performing the altruistic act increases the image score of the donor and may therefore increase the chance of obtaining a benefit in a future encounter as a recipient. On the other hand, a discriminator who punishes low-score players by refusing them help pays for this by having his own score reduced. The overriding idea, relevant to human societies, is that information about another player does not require a direct interaction, but can be obtained indirectly either by observing the player or by talking to others. The evolution of human language as a means of such information transfer has certainly helped in the emergence of cooperation based on indirect reciprocity. □

Methods

Defectors and discriminators. Here we develop a simplified model for indirect reciprocity which can be fully understood in analytical terms. Consider two image scores, 0 for someone who defected last round and 1 for someone who cooperated last round. Thus the image score depends only on the last move of a player as a donor. Consider two types of players: discriminators, who help only players with image score 1, and defectors, who never help. Let us suppose that there is a probability, q , that discriminators have information about the image score of the recipient. In the absence of information, they assume an image score of 1 with probability p . (One can show that if indirect reciprocity works at all, then discriminators with larger p always outcompete the others²⁸. Therefore we shall restrict ourselves in the following to the limiting value $p = 1$. The discriminator strategy in this case coincides with a variant of 'tit-for-tat', which begins with defection if the future co-player has been seen defecting in his last interaction²⁹—a confirmation of Alexander's view that "indirect reciprocity is a consequence of direct reciprocity occurring in the presence of others"¹⁷.) For a defector, information about the image score does not matter. We denote by x_0, x_1, y_0 and y_1 , respectively, the frequencies of discriminators (x) with images 0 and 1, and the frequencies of defectors (y) with images 0 and 1. The total frequency of discriminators is $x = x_0 + x_1$ and that of defectors is $y = y_0 + y_1$. We have $x + y = 1$. A generation consists of several rounds of the game, during which x and y do not change. In each round all players are paired up, half of the players being donors, the other half recipients. The frequencies of players of image 0 or 1 change from round to round according to the difference equations $x'_0 = [x_0 + x(1-\psi)q]/2$, $x'_1 = [x_1 + x(1-q+q\psi)]/2$, $y'_0 = [y_0 + y]/2$, and $y'_1 = y_1/2$. Here $\psi = x_1 + y_1$ is the frequency of players with score 1. In each round, the payoff to the

individual types is $P(x_0) = [-c(1-q+q\psi) + bx(1-q)]/2$, $P(x_1) = [-c(1-q+q\psi) + bx]/2$, $P(y_0) = bx(1-q)/2$, $P(y_1) = bx/2$. The difference equation yields the expected payoff values $D_e(k)$ and $D_i(k)$ to defectors and discriminators in the k th round: $D_e(k) = bx(1-q+q2^{-(k-1)})/2$, and $D_i(k) = D_e(k) + \{(1-q)(bqx - c)(1-qx)^{-1} - bq2^{-(k-1)} + q(b-c)(1-x)(1-qx)^{-1}[(1+qx)/2]^{k-1}\}/2$. We can assume either that the number of rounds per generation is constant, or that there exists a fixed probability w for a further round. In the latter case, the total payoff to defectors is $D_e = \sum_{k=1}^{\infty} w^{k-1} D_e(k)$, and similarly for discriminators. We find that

$$2(D_i - D_e) = (1-q) \frac{bqx - c}{(1-w)(1-qx)} + 2q \left[\frac{(b-c)(1-x)}{(1-qx)(2-w-wqx)} - \frac{b}{2-w} \right]$$

Modelling the change in frequency of discriminators and defectors from one generation to the next by the standard replicator equation³⁰, we find that defectors win if x is below a threshold value $x_{\min}$ given by $D_i = D_e$, whereas discriminators win if x is above this threshold. Discriminators are evolutionarily stable if and only if $x_{\min} < 1$, that is, if $D_i > D_e$ for $x = 1$. This can happen only for $q > c/b$, that is, if the probability of knowing the image of the co-player exceeds the cost-to-benefit ratio, and if the average number of rounds, that is, $1/(1-w)$, exceeds $(bq+c)/(bq-c)$. Note that for our numerical example of Figs 1 and 2, where $b = 1$, $c = 0.1$ and $q = 1$, we need only about 1.2 rounds per generation for cooperation to be stable against invasion by defectors.

The good, the bad and the discriminating. Indirect reciprocity only works when donors discriminate between individuals that have or have not helped others in the past. To understand the role of indiscriminate altruists, we add to the population of defectors and discriminators a fraction z of cooperators, who always give help irrespective of the their co-player's score. We can calculate the payoffs in each round as before. The cooperators' total expected payoff, D_c , differs from that of the defectors, D_e , by $[-c + (bwqx)/(2-w)]/[2(1-w)]$, whereas

$$D_i - D_e = \frac{(bqx - c)(1 - q + qz)}{2(1 - w)(1 - qx)} - \frac{bq(x + y)}{2 - w} + \frac{qy(b - c)}{(1 - qx)(2 - w - wqx)}$$

The population is in equilibrium whenever $y = 0$ (no defectors) or $x = c(2-w)/bwq$. If x lies below the latter value, defectors win; if x exceeds it, then a mixture of discriminating and indiscriminating altruists gets established, depending on the initial value. This mixed state is proof against invasion by unconditional defectors, but in such a population both discriminating and indiscriminating altruists do equally well. Their frequencies will be altered only by random drift, not by selection. If the frequency of discriminators falls below $c(2-w)/bwq$ then defectors can invade and take over. Defectors in turn can be overcome by discriminators if their frequency fluctuates above $x_{\min}$.

A universal constant of nature. Let us now consider a situation in which the image score can be any integer number between $-\infty$ and $+\infty$, but in which all players adopt the same strategy, $k = 0$. Denote by x_i the frequency of players with image score i . In the next round it is $x'_i = [x_i + x_{i-1}\phi + x_{i+1}(1-\phi)]/2$ where $\phi = \sum_{i=0}^{\infty} x_i$. If all players start with an image score of greater than or equal to 0, then all players will cooperate in the first and all subsequent rounds. If all players start with an image score of less than 0, then all players will defect in the first and all subsequent rounds. The situation becomes interesting if there is an initial distribution of image scores above and below 0. The question of whether the system will ultimately converge to cooperation or defection is non-trivial. We find that there is a maximum fraction of players with an initial image score below 0, such that the system ultimately converges to all-out cooperation. Numerical simulations show that this fraction is 0.7380294688360....

Received 11 November 1997; accepted 31 March 1998.

- Hamilton, W. D. The evolution of altruistic behaviour. *Am. Nat.* **97**, 354–356 (1963).
- Hamilton, W. D. The genetical evolution of social behaviour. *J. Theor. Biol.* **7**, 1–16 (1964).
- Williams, G. C. *Group Selection* (Aldine-Atherton, Chicago, 1971).
- Eshel, I. On the neighbourhood effect and evolution of altruistic traits. *Theor. Popul. Biol.* **3**, 258–277 (1972).
- Wilson, D. S. & Sober, E. Reintroducing group selection to the human behavioural sciences. *Behav. Brain Sci.* **17**, 585–654 (1994).
- Trivers, R. The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57 (1971).
- Axelrod, R. & Hamilton, W. D. The evolution of cooperation. *Science* **211**, 1390 (1981).
- Axelrod, R. *The Evolution of Cooperation* (Basic Books, New York, 1984).
- Nowak, M. A. & May, R. M. Evolutionary games and spatial chaos. *Nature* **359**, 826–829 (1992).

10. Michod, R. E. & Sanderson, M. J. in *Evolution: Essays in Honor of John Maynard Smith* (eds Greenwood, P. J., Harvey, P. & Slatkin, M.) 95–106 (Cambridge Univ. Press, Cambridge, 1985).
11. Peck, J. & Feldman, M. The evolution of helping in large, randomly mixed populations. *Am. Nat.* **127**, 209–221 (1985).
12. Milinski, M. Tit for tat in sticklebacks and the evolution of cooperation. *Nature* **325**, 433–435 (1987).
13. May, R. M. More evolution of cooperation. *Nature* **327**, 15–17 (1987).
14. Dugatkin, L. A. Mesterton-Gibbons, M. & Houston, A. I. Beyond the prisoner's dilemma: towards models to discriminate among mechanism of cooperation in nature. *Trends Ecol. Evol.* **7**, 202–205 (1992).
15. Nowak, M. A. & Sigmund, K. Tit for tat in heterogeneous populations. *Nature* **355**, 250–253 (1992).
16. Nowak, M. A. & Sigmund, K. Win–stay, lose–shift outperforms tit for tat. *Nature* **364**, 56–58 (1993).
17. Alexander, R. D. *The Biology of Moral Systems* (Aldine de Gruyter, New York, 1987).
18. Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge Univ. Press, Cambridge, 1982).
19. Wilson, E. O. *Sociobiology* (Harvard Univ. Press, Cambridge, MA, 1975).
20. Krebs, J. R. & Davies, N. B. *An Introduction to Behavioural Ecology* (Blackwell, Oxford, 1987).
21. Buss, L. *The Evolution of Individuality* (Princeton Univ. Press, NJ, 1987).
22. Frank, S. A. The origin of synergistic symbiosis. *J. Theor. Biol.* **176**, 403–410 (1995).
23. Binmore, K. G. *Fun and Games: a Text on Game Theory* (Heath, Lexington, MA, 1992).
24. Marler, P. & Evans, C. Bird calls: just emotional displays or something more? *Ibis* **138**, 26–33 (1996).
25. Zahavi, A. & Zahavi, A. *The Handicap Principle: A Missing Piece of Darwin's Puzzle* (Oxford Univ. Press, Oxford, 1997).
26. Boyd, R. & Richerson, P. J. The evolution of indirect reciprocity. *Social Networks* **11**, 213–236 (1989).
27. Sugden, R. *The Economics of Rights, Cooperation and Welfare* (Blackwell, Oxford, 1986).
28. Nowak, M. A. & Sigmund, K. The dynamics of indirect reciprocity. *J. Theor. Biol.* (submitted).
29. Pollock, G. B. & Dugatkin, L. A. Reciprocity and the evolution of reputation. *J. Theor. Biol.* **159**, 25–37 (1992).
30. Hofbauer, J. & Sigmund, K. *Evolutionary Games and Population Dynamics* (Cambridge Univ. Press, Cambridge, 1998).

Acknowledgements. We thank M. Dawkins, A. Kacelnik, J. Krebs and R. May for discussion. Support from the Wellcome Trust is gratefully acknowledged. Part of this work was done at IASA (Laxenburg).

Correspondence and requests for materials should be addressed to M.A.N. (e-mail: martin.nowak@zoo.ox.ac.uk).

Selective representation of relevant information by neurons in the primate prefrontal cortex

Gregor Rainer, Wael F. Asaad & Earl K. Miller

Department of Brain and Cognitive Sciences and The Center for Learning and Memory, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

The severe limitation of the capacity of working memory, the ability to store temporarily and manipulate information¹, necessitates mechanisms that restrict access to it. Here we report tests to discover whether the activity of neurons in the prefrontal (PF) cortex, the putative neural correlate of working memory^{2–8}, might reflect these mechanisms and preferentially represent behaviourally relevant information. Monkeys performed a 'delayed-matching-to-sample' task with an array of three objects. Only one of the objects in the array was relevant for task performance and the monkeys needed to find that object (the target) and remember its location. For many PF neurons, activity to physically identical arrays varied with the target location; the location of the non-target objects had little or no influence on activity. Information about the target location was present in activity as early as 140 ms after array onset. Also, information about which object was the target was reflected in the sustained activity of many PF neurons. These results suggest that the prefrontal cortex is involved in selecting and maintaining behaviourally relevant information.

In the 'array trials', a sample array of three objects was briefly presented while the monkeys maintained central gaze (Fig. 1a). Monkeys needed to find the target object in the array and remember its location. After a brief delay, a test array appeared and the monkeys had to release a lever if the target object appeared in the same location as it had in the sample array. Although each of the three objects was a target, in turn, for a block of trials, its location in the sample array was chosen randomly on each trial. Monkeys were cued to the target object with 'cue trials' (Fig. 1a) in which the target object appeared alone.

Table 1 Summary of neuronal selectivity in different task periods

	Sample period	Delay period	Both periods
<i>n</i> = 97 cells			
Number of cells selective for:			
Target object only	20	16	7
Target location only	22	30	13
Target object and location	24	15	8
Total selective for object	44	31	15
Total selective for location	46	45	21
Selectivity depth:			
Object	48%	44%	–
Location	48%	53%	–
Selectivity index:			
Object	0.24	0.24	–
Location	0.24	0.28	–

Cell counts are based on ANOVA (see Methods), evaluated at $P < 0.01$. Mean selectivity depths and selectivity indices were computed from delay activity on array trials for cells showing a significant ANOVA. For cells not showing significant effects, mean selectivity depths ranged from 12 to 15% and mean selectivity indices ranged from 0.08 to 0.09.

We recorded the activity of 97 neurons from the lateral prefrontal cortex of two monkeys (Fig. 1b). Based on analysis of variance (ANOVAs) (evaluated at $P < 0.01$), many PF neurons showed activity to physically identical sample arrays that varied depending on which of three array positions contained the target (46/97 or 47% during sample presentation, 45/97 or 46% during the delay, Fig. 2 and Table 1). Information about the target location appeared very early in neural activity, starting about 140 ms after array onset (Fig. 3a). The activity of these neurons after this time largely reflected the target location alone; information about the location of the irrelevant, non-target, objects had little or no influence. Although almost half of PF cells showed activity that varied with the location of the target object, only a few cells (sample period: 10/97 or 10%; delay period: 5/97 or 5%) showed activity that varied with the location of non-target objects (t -tests, evaluated at $P < 0.01$). In fact, many cells showed similar activity on array trials and on cue trials in which the target object appeared alone (Fig. 2).

The task also required monkeys to remember which object was currently the target. This was also reflected in PF activity. On array trials, many PF neurons showed activity during the sample period

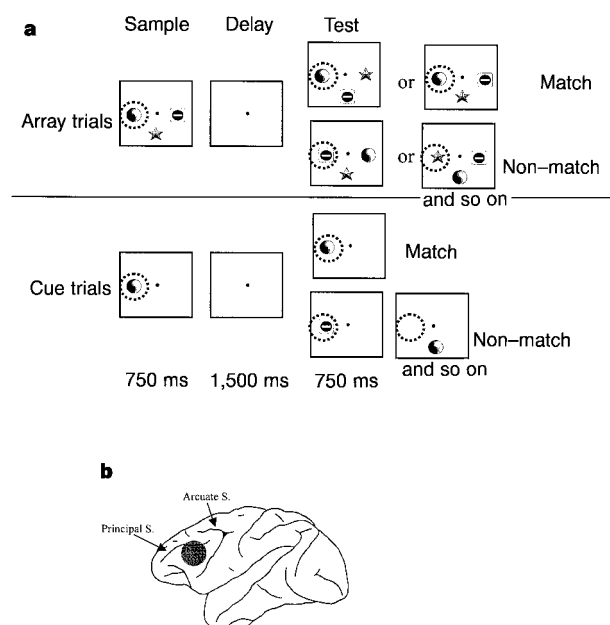


Figure 1 The behavioural task and recording sites. **a**, Sequence of trial events. Each trial began when the monkey grasped a lever and fixated a small fixation target at the centre of a computer screen. The location of the target object is indicated by the dotted circle on this figure. Examples of array trials (top) and cue trials (bottom) are illustrated. **b**, Location of recording sites: Arcuate S, arcuate sulcus; Principal S, principal sulcus.

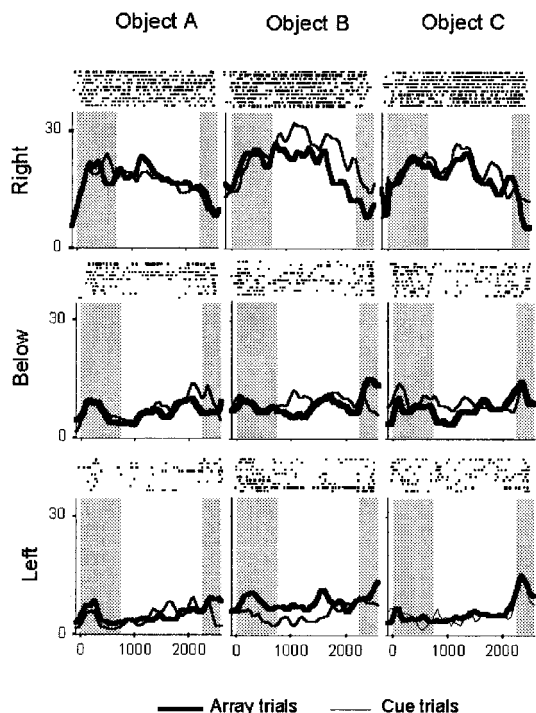


Figure 2 Delay activity from a single PF neuron that varied with the location of the target object. The grey bar on the left of each histogram indicates time of sample presentation and the grey bar on the right indicates presentation of the test array. The column labels refer to the object that was the target in the array trials or which single object was used as the sample on cue trials. The row labels refer to the location of the target. Bin width, 40 ms. Above each histogram are rasters from 10 trials of array presentation. Each dot represents an action potential from the neuron and each row corresponds to a different trial.

(44/97 or 45%) and during the delay (31/97 or 32%) that varied depending on which object was relevant. However, unlike activity selective for the target location, information about the target object was reflected in activity even before the appearance of the sample array on each trial (Fig. 5b). Because a given object was the target for an entire block of trials, this chronic change in activity presumably reflects its maintained memory across trials.

These results suggested that monkeys were focusing their attention on the target location. In behavioural experiments, we confirmed that the target was 'capturing' the attention of the monkeys. The monkeys were allowed to look freely at the sample array. Under free gaze, eye movements and attention are closely coupled^{9,10}. During the 750 ms of the sample array, the monkeys spent, on average, 529 ms looking at the target and only 82 ms looking at each of the non-targets. Thus, the requirement to remember the target resulted in the monkeys directing attention to it.

It is well established that working memory is severely limited in capacity^{11–13}. Our results suggest that focal attention can play a major role in regulating access to working memory. Information about the location of an attended target dominated the activity of many PF neurons, the putative neural correlate of working memory. Thus, we hold in working memory that to which we attend. But the converse is also true. That is, working memory can direct attention: PF activity provided a representation of the to-be-attended target object that could have been used to select it from the array¹⁴. Psychological studies and models of selective attention have suggested this intimate relationship between attention and working memory^{14,15}. These results illustrate it in prefrontal activity and provide support for the notion that the attentional effects observed in regions such as the parietal cortex¹⁶ and the inferior temporal cortex¹⁷ arise from bias signals originating in the PF cortex¹⁵. □

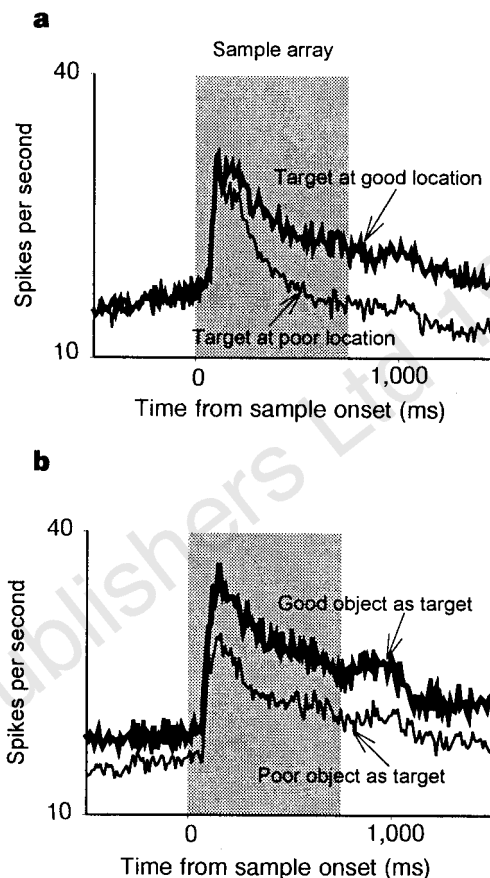


Figure 3 Time course of location and object effects for cells showing those effects. **a**, Average activity of 45 PF neurons selective for the target location. **b**, Average activity of 31 PF neurons selective for the target object. Neurons were selected for this figure if they showed a significant effect in the last 1,000 ms of the 1,500 ms delay. The grey bar shows the time of sample array presentation. The difference in activity when a good object was the target and when a poor object was the target was evident in the baseline activity before sample onset (t -test, $P < 0.001$).

Methods

Behavioural task. The task is illustrated in Fig. 1. The objects were colour pictures of objects 2° by 2° in size. 'Real world' stimuli such as these have been shown to elicit selective responses readily from prefrontal neurons⁶. For each day of recording, three objects were chosen at random from a large pool of objects. We did not determine which features of the object elicited activity from the neurons under study; for the purposes of this study all that mattered was whether different objects or their locations elicited different levels of activity.

Monkeys were required to maintain central gaze throughout the trial. Each of the objects appeared 4° to the right, to the left, and below fixation. The location of each of the three objects in the sample array was chosen randomly on each trial. Each object was the target, in turn, for a block of about 80 trials. Each block began with 10 cue trials in which the target object was used alone as the sample. The cue trials were then randomly intermixed with array trials throughout the blocks. Blocks in which a given object was the target were interleaved with blocks with another target object until there were about 9–12 blocks (3–4 repetitions of each block of each object as a target). This interleaved design compensates for any changes in neural activity across blocks that were unrelated to the task. Monkeys were well trained for this task, averaging over 85% correct.

Electrophysiological recording. Recordings were made in the lateral prefrontal cortex (see Fig. 1b) of two adult rhesus monkeys (*Macaca mulatta*) using standard electrophysiological techniques. Recording sites were

localized using magnetic resonance imaging. We advanced the electrode until neuronal activity was well isolated. Then, data collection commenced. To ensure an unbiased assessment of PF activity, we made no attempt to select neurons based on task-related activity.

Data analysis. Visual responses to the samples were analysed over an interval from 100 ms to 750 ms after sample onset. Delay activity was analysed over the last 1,000 ms of the 1,500 ms delay after the sample array.

Activity was appraised using ANOVA, evaluated at $P < 0.01$. To determine whether activity on array trials reflected the target object, its location, or both, a two-factor ANOVA was used. One factor was which object was the target and the other was its location. To assess the influence of the location of non-target objects, we computed nine t -tests for each cell (evaluated at $P < 0.01$, with Bonferroni correction for multiple tests). Each t -test compared responses to the two possible arrays that contained a given target at a given location. For example, one array might contain object A as a target at position 1, B at position 2, and C at position 3 (that is, A1 B2 C3) whereas the other would contain target A at 1, C at 2, and B at 3 (A1 C2 B3). Thus, the arrays differed in the position of the non-target objects only. This allowed us to determine whether activity to an array with a given target (such as A) at a given location (such as 1) was influenced by the positions of the two non-targets (such as B and C).

'Good' and 'poor' refer to the object or location that elicited the most or least activity, respectively. Selectivity depth measured the difference in activity on array trials when a good versus poor object or location was relevant. It was computed by dividing the difference in their activity by their sum and then converting this value to per cent difference¹⁸. The selectivity index¹⁹, by contrast, takes into account changes in activity to each of the three relevant objects or locations. It was computed on array trial activity and was defined as:

$$S = \frac{n - \left(\frac{\sum r_i}{r_{\max}} \right)}{n - 1}$$

n = number of objects or locations; r_i = activity to a target object or location i ; $r_{\max}$ is the maximum r_i . A value of zero indicates identical responses to all stimuli; a value of 1 indicates activation by one stimulus and silence to other stimuli. Values typically do not approach 1 because PF neurons are rarely silent and have some baseline firing.

Received 1 December 1997; accepted 25 March 1998.

1. Baddeley, A. *Working Memory* (Clarendon, Oxford, 1986).
2. Fuster, J. M. & Alexander, G. E. Neuron activity related to short-term memory. *Science* **173**, 652–654 (1971).
3. Kubota, K. & Niki, H. Prefrontal cortical unit activity and delayed alternation performance in monkeys. *J. Neurophysiol.* **34**, 337–347 (1971).
4. Funahashi, S., Bruce, C. J. & Goldman-Rakic, P. S. Mnemonic coding of visual space in the monkey's dorsolateral prefrontal cortex. *J. Neurophysiol.* **61**, 331–349 (1989).
5. di Pellegrino, G. & Wise, S. P. A neurophysiological comparison of three distinct regions of the primate frontal lobe. *Brain* **114**, 951–978 (1991).
6. Miller, E. K., Erickson, C. A. & Desimone, R. Neural mechanisms of visual working memory in prefrontal cortex of the macaque. *J. Neurosci.* **16**, 5154–5167 (1996).
7. Rao, S. C., Rainer, G. & Miller, E. K. Integration of what and where in the primate prefrontal cortex. *Science* **276**, 821–824 (1997).
8. Hasegawa, R., Sawaguchi, T. & Kubota, K. Monkey prefrontal neuronal activity coding the forthcoming saccade in an oculomotor delayed matching to sample task. *J. Neurophysiol.* **79**, 322–333 (1998).
9. Hoffman, J. E. & Subramaniam, B. Saccadic eye movement and visual selective attention. *Percept. Psychophys.* **57**, 787–795 (1995).
10. Shepard, M., Findlay, J. M. & Hockley, R. J. The relationship between eye movements and spatial attention. *Q. J. Exp. Psychol.* **38A**, 475–491 (1997).
11. Miller, G. A. The magic number seven plus or minus two: some limits on capacity for processing information. *Psychol. Rev.* **63**, 81–97 (1956).
12. Kahneman, D. *Attention and Effort* (Prentice-Hall, Englewood Cliffs, New Jersey, 1973).
13. Luck, S. J. & Vogel, E. K. The capacity of visual working memory for features and conjunctions. *Nature* **390**, 279–281 (1997).
14. Duncan, J. & Humphreys, G. W. Visual search and similarity. *Psychol. Rev.* **96**, 433–458 (1989).
15. Desimone, R. & Duncan, J. Neural mechanisms of selective visual attention. *Annu. Rev. Neurosci.* **18**, 193–222 (1995).
16. Gottlieb, J. P., Kusunoki, M. & Goldberg, M. E. The representation of visual salience in monkey parietal cortex. *Nature* **391**, 481–484 (1998).
17. Chelazzi, L., Miller, E. K., Duncan, J. & Desimone, R. A neural basis for visual search in inferior temporal (IT) cortex. *Nature* **363**, 345–347 (1993).
18. Luck, S. J., Chelazzi, L., Hillyard, S. A. & Desimone, R. Neural mechanisms of spatial selective attention in areas V1, V2, and V4 of macaque visual cortex. *J. Neurophysiol.* **77**, 24–42 (1997).
19. Moody, S. L., Wise, S. P., Di Pellegrino, G. & Zipser, D. A model that accounts for activity in primate frontal cortex during a delayed matching-to-sample task. *J. Neurosci.* **18**, 399–410 (1998).

Acknowledgements. We thank S. Chenchal Rao for his participation in an early phase of this experiment and M. Histed for expert animal training and computer support. This work was supported by an NINDS grant and the Pew Charitable Trusts. We thank P. Dayan, R. Desimone, S. Macknik, J. Mazer, J. Schall, R. Webb, M. Wickersham and M. Wilson for their valuable comments.

Correspondence and requests for materials to E.K.M. (e-mail: ekm@ai.mit.edu).

Gli/Zic factors pattern the neural plate by defining domains of cell differentiation

R. Brewster, J. Lee & A. Ruiz i Altaba

The Skirball Institute, Developmental Genetics Program, Department of Cell Biology, NYU Medical Center, 540 First Avenue, New York, New York 10016, USA

Three cell types differentiate in the early frog neural plate: neural crest at the lateral edges, floorplate at the midline and primary neurons in three bilateral stripes. Floorplate cells and ventral neurons are induced by Sonic hedgehog^{1,2} (Shh) and neural crest and dorsal neurons are induced by epidermal factors such as bone morphogenetic proteins (BMPs)³. Neurogenesis in a subset of cells within the stripes involves lateral inhibition⁴. However, the process by which pools of precursors are defined in stereotypic domains in response to inductive signals is unknown. Here we show that frog *Zic2* encodes a zinc-finger transcription factor of the Gli superfamily which is expressed in stripes that alternate with those in which primary neurons differentiate and overlap the domains of floorplate and neural crest progenitors. *Zic2* inhibits neurogenesis and induces neural crest differentiation. Conversely, Gli proteins are widely expressed, induce neurogenesis and inhibit neural crest differentiation. *Zic2* is therefore a vertebrate pre-pattern gene, encoding anti-neurogenic and crest-inducing functions that counteract the neurogenic but not the floorplate-inducing activity of Gli proteins. We propose that the combined function of Gli/Zic genes responds to inductive signals and induces patterned neural cell differentiation.

Frog *Zic2* encodes a nuclear protein (Fig. 1a, inset) homologous to mouse *Zic2* (refs 5, 6) and all *Zic* proteins seem to recognize Gli binding sites⁷. *Zic2* messenger RNA is detected in the dorsal ectoderm of the early gastrulae and early neurulae (Fig. 1a). *Zic2* is then expressed in stripes alternating and non-overlapping with those of primary *N-tubulin*⁺ (*N-tub*⁺) neurons, although not all *N-tub*-negative areas express *Zic2* (Figs 1b, c, e and 2a, b). Midline and, later, floorplate cells show low levels of *Zic2* expression (Fig. 1d, top). The most medial stripe of *N-tub*⁺ primary motor neurons is located in between the midline, which expresses *Shh*², and a medial *Zic2* stripe (Fig. 1d, e). The intermediate stripes of *N-tub*⁺ interneurons in anterior regions is in between the medial *Zic2* stripe and its expression in the neural folds (Figs 1b, c and 2a, b). Posteriorly, the outer stripe of *N-tub*⁺ neurons is lateral to *Zic2* in the neural folds (Figs 1b–d and 2a, b). Trigeminal ganglion neurons are also adjacent to the anterior domain of *Zic2* (Fig. 2a, inset). At later stages, *Zic2* is expressed at high levels in the anterior edge, the future anterior brain being marked by delayed *N-tub* expression (Fig. 1b, c). *Zic2* is also detected at high levels in the lateral neural folds, which contain the presumptive cranial neural crest (Fig. 1b), and later in early migrating hindbrain crest (Fig. 1f). In tadpoles, *Zic2* is expressed in distinct regions of the brain (Fig. 1f, g) and in the spinal cord, it is detected dorsally (Fig. 1g).

To test whether *Zic2* participates in defining neurogenic domains, we analysed the expression patterns of *N-tub* (Fig. 2b) and *Neurogenin1* (ref. 8) (*Ngnr1*; Fig. 2e) in injected embryos. Unilateral *Zic2* misexpression resulted in the unilateral loss of *N-tub*⁺ and *Ngnr1*⁺ cells (Fig. 2c, f) that inherited *Myc*–*Zic2* protein or β -galactosidase from injected RNAs. Suppression of neurogenesis is not due to inhibition of neural development, as the neural plate of most of the injected embryos displayed normal expression of neural cell adhesion molecule (NCAM) (Fig. 2d: stage ~14, 66% normal, $n = 18$; and data not shown). We also analysed the ability of *Zic2* to inhibit ectopic neurogenesis, downstream of neural induction,

driven by exogenous *Ngnr1* protein. Misexpression of *Ngnr1* (ref. 8) or other neurogenic basic-helix–loop–helix (bHLH) proteins (see, for example, ref. 9) results in the robust induction of neuronal differentiation throughout the neural plate and adjacent ectoderm (Fig. 2g). In contrast, co-injection of equal amounts of *Ngnr1* and *Zic2* caused a drastic decrease in the number of ectopic neurons (Fig. 2h) and occasional loss of endogenous neurons.

Zic2 may be a repressor as it has mono-amino-acid stretches characteristic of repressor domains¹⁰. We found that expression of the strong transactivating domain of adenovirus VP16 fused to the zinc-finger domain of *Zic2* (VP16–*Zic2*) resulted in a phenotype opposite to that of *Zic2*: ectopic neurogenesis and ectopic expression of *Ngnr1* (Fig. 2i–k). These results indicate that *Zic2* normally represses neurogenic factors.

VP16–*Zic2* could mimic other neurogenic Gli superfamily members as they all bind the same ‘Gli’ target sites *in vitro*^{7,11}. Indeed, *Gli1* induces neuronal differentiation¹² but its restricted expression precludes its involvement in the generation of most neurons. In contrast, *Gli2* and *Gli3* are good candidates for the induction of neuronal differentiation as they are expressed throughout most of the neural plate^{12,13}. Embryos injected with frog *Gli1*, *Gli2* or *Gli3*, but not control embryos, displayed concentration-dependent ectopic neurogenesis (Fig. 3a–d, h, j), strongest with *Gli2* and weakest with *Gli3*. The differentiation of ectopic neurons induced by *Gli1* was accompanied by the repression of sensory neurons of the outer stripe in a fraction of embryos (Fig. 3h). Ectopic differentiation of induced neurons by injected Gli proteins was preceded by the ectopic expression of *Ngnr1* (results not shown; 53% ectopic, *n* = 51 for *Gli1*; 21%, *n* = 39 for *Gli2*; 22%, *n* = 51 for *Gli3*) and confirmed at tadpole stages with the pan-neural marker *Xen1* (ref. 14, Fig. 3e).

To test whether Gli proteins are involved in activating endogenous neurogenesis, we fused the repressor domain of the *Drosophila*

Engrailed protein¹⁰ to the DNA-binding zinc-finger domain of *Gli2* to create EnR–*Gli2*. We chose *Gli2* because it is expressed widely and has strong neurogenic activity. Expression of EnR–*Gli2* in the neural plate resulted in the absence of *N-tub*⁺ neurons, which was not due to inhibition of neural development (95% inhibition of neurogenesis (Fig. 3f, g) versus 37% lower NCAM expression; *n* = 26; results not shown), although at high levels EnR–*Gli2* may interfere with neural-inducing proteins of the Gli superfamily. EnR–*Gli2* also inhibited the ectopic differentiation of *N-tub*⁺ neurons induced by co-injected *Gli2* (Fig. 3j, k). To test whether EnR–*Gli2* could inhibit other Gli proteins, we assayed for ectopic HNF-3 β ⁺ floorplate induction by *Gli1*. Alone, *Gli1* induced ectopic floorplate cells¹² (50% ectopic; *n* = 14; results not shown), whereas *Gli1* and co-injected EnR–*Gli2* did not (6% ectopic; *n* = 32; not shown). Together, these results indicate that endogenous neurogenesis and ectopic floorplate differentiation require Gli protein function.

We then tested for a possible repression of Gli function by *Zic2*. Embryos co-injected with *Zic2* plus *Gli1* or *Gli2* showed a marked decrease in ectopic neurogenesis (Fig. 3i, l). *Zic2* mediated strong repression of ectopic neurogenesis by *Gli1* and a small fraction of injected embryos also had repressed endogenous neurogenesis (Fig. 3i). The repression of *Gli2* function by equal amounts of *Zic2* was incomplete, consistent with the more potent neurogenetic function of *Gli2* versus *Gli1* (Fig. 3l). *Zic2* is expressed transiently in the floorplate, but injected embryos did not show defects in endogenous or ectopic floorplate differentiation (*n* = 44; not shown). Co-injected *Zic2* did not inhibit ectopic induction of HNF-3 β ⁺ floorplate cells by *Gli1* either (80% ectopic for *Gli1* alone, *n* = 15; 81% for *Gli1* plus *Zic2*, *n* = 16; results not shown). Thus *Zic2* is an inhibitor of neuronal but not floorplate differentiation.

Anterior *Zic2* expression is detected adjacent to and partially overlapping that of the neural crest marker *Slug*¹⁵ (Fig. 4a, and

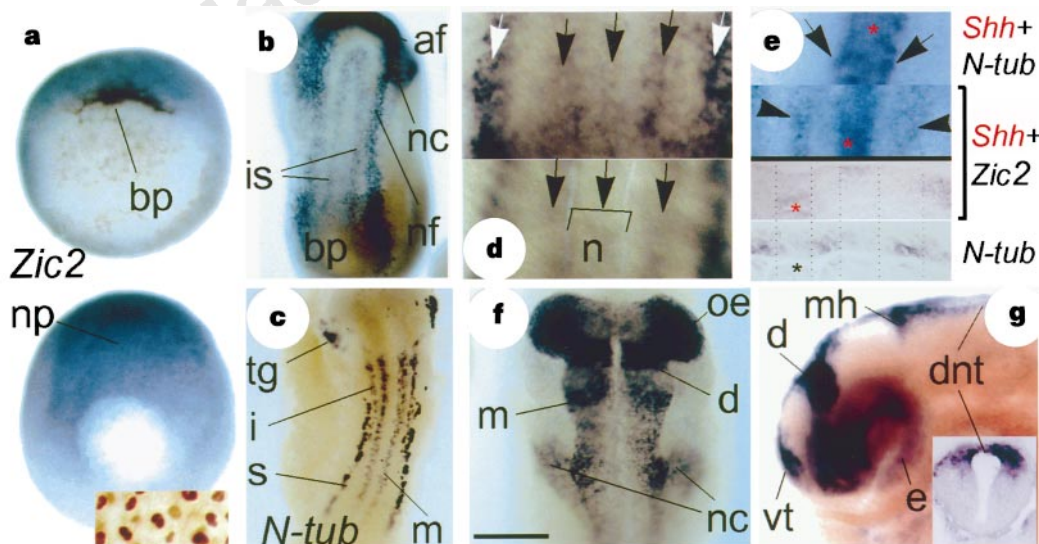


Figure 1 Neural expression of *Zic2* in frog embryos. **a**, Expression of *Zic2* mRNA in the neural plate (np) (top, stage 10+; bottom, embryonic stage ~11.5) adjacent to the blastopore (bp). Inset, nuclear localization of Myc-tagged *Zic2* protein (stage ~11–12). **b**, Localization of *Zic2* in the neural folds (nf) (stage ~14), in internal stripes (is) and the cranial neural crest (nc) lateral to the anterior neural fold (af). **c**, Primary neurons in neurulae (stage ~15/16) are marked by *N-tub* in interneurons (i), motor (m), sensory (s) and trigeminal ganglion (tg) neurons. **d**, Expression of *Zic2* (arrows) in bilateral stripes, in the neural plate edges and in midline cells at the focal plane of the neural plate (top) or of the underlying notochord (n, bracketed, bottom). Views of the posterior-most area (stage 14/15) are shown. **e**, Whole-mount views (top two panels and histological cross-sections (bottom two panels) showing expression of *Shh* (red asterisks) in midline cells and *N-tub* or *Zic2* (black arrows, top). Note the position of primary motor neurons, flanking the midline

overlying the notochord (black asterisk in bottom panel), located in between the *Shh*⁺ midline and the internal *Zic2* stripes. The alternating domains of *Ntub* and *Zic2* are delimited in one side of the neural plate by dashed lines (two bottom panels). The same chromogen was used for all probes. Stage ~14. **f**, *Zic2* in tailbud (stage 20–22) embryos is absent from the midline but shows boundaries in the optic evagination (oe), diencephalon (d) and midbrain (m). Expression is also detected in the dorsal neural tube and hindbrain crest (nc). **g**, *Zic2* in zones of the tadpole brain (stage ~34). Expression is detected in the dorsal neural tube (dnt; see inset), the dorsal midbrain/hindbrain boundary (mh), the diencephalon (d), ventral telencephalon (vt) and eye (e). All panels show dorsal views, except for **a** and **g**, which show vegetal and lateral views. Scale bar, 340 μ m for **a**; 230 μ m for **b**, **c**; 100 μ m for **d**, **e**; and 180 μ m for **f**, **g**.

results not shown), suggesting a role for *Zic2* in crest development. Injection of *Zic2* resulted in ectopic expression of *Slug* and *Snail*¹⁶, two cranial crest markers, and *Livertine*¹⁷, a general crest/dorsal neural tube marker, in non-neural ectoderm (Fig. 4a–g). Ectopic *Snail* and *Slug* were often restricted to the anterior half of the embryo, suggesting the existence of global anterior–posterior (A–P) pattern. *Livertine* labelling showed induction of ectopic crest throughout the axis (Fig. 4g, and results now shown) in the absence of ectopic general neural (*Xen1*, $n = 117$; *Xen2*, $n = 55$; or *Xen3*, $n = 29$; ref. 14; results now shown) or lateral neural plate/dorsal neural tube (*Pax3*; Fig. 4h, i; or *Gli3*, $n = 26$; not shown) markers. Expression of VP16–*Zic2* or Gli proteins resulted in the opposite phenotype, yielding an inhibition of *Slug* expression (Fig. 4j, k). This is consistent with the repression of cranial crest differentiation by an activating *Gli3* chimera¹³. *Zic2* may therefore normally induce and *Gli2/3* restrict the domain of cranial crest differentiation.

We propose that following neural induction, widespread expression of *Gli2* and *Gli3* together with localized expression of *Gli1* and

Zic2 define domains of neural cell differentiation. The neural- and crest-inducing activities of *Zic1* (ref. 18) and *Zic3* (ref. 19) probably also participate. Our results also suggest that the differentiation of distinct neuronal types depends on local Gli combinations²⁰. The neurogenic function of Gli proteins and induction of *Gli1* by *Shh*¹² is analogous to induction of *atonal* by Hedgehog via Cubitus interruptus²¹, three *Drosophila* homologues of *Ngnr1*, *Shh* and *Gli*, respectively. *Zic2* may likewise act upstream of the bHLH neurogenic cascade as it represses endogenous transcription of *Ngnr1*, a key early neurogenic regulator⁸. However, it also inhibits *Ngnr1* activity in co-injection assays. *Zic2* may act indirectly, perhaps through antineurogenic HES proteins²², or directly by repressing neurogenic bHLH gene expression and function. *Zic2* thus acts as one of several possible pre-pattern genes delimiting where and when primary neuronal precursors can differentiate. Indeed, injected *Zic2* only seems to delay neurogenesis, as low levels of tagged protein can be detected in the neural tube (results not shown). In brain, *Zic2* may also be involved in the definition of

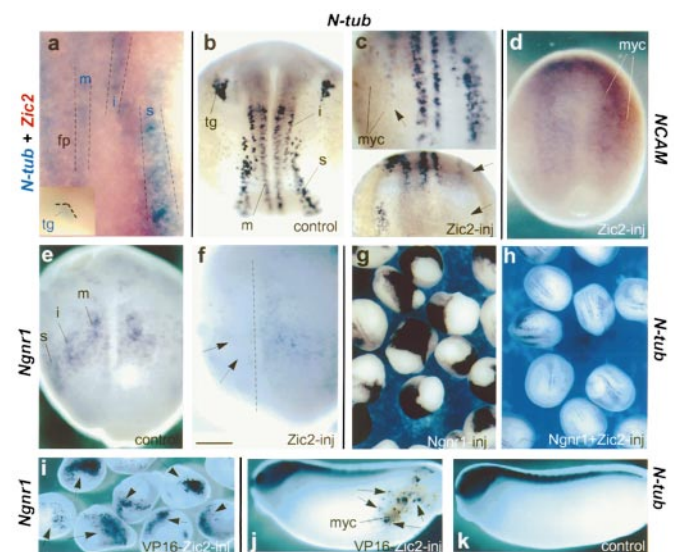


Figure 2 Effects of *Zic2* and VP16–*Zic2* on neurogenesis. **a**, Detail of a double *in situ* hybridization showing *Ntub*⁺ primary neurons (blue) adjacent to *Zic2*⁺ regions (pink). The low levels of expression of *Zic2* in the internal stripes make their distinction difficult. Inset, non-overlap of *Zic2* and *Ntub*⁺ in the trigeminal ganglion (tg). Stage ~14. **b**, **e**, *Ntub* (**b**; stage ~14; $n = 19$) and *Ngnr1* (**e**; stage ~11; $n = 31$) in control embryos. The punctate labelling in **b** is weak *Ntub* expression is neuromasts. **c**, **f**, Absence of expression of *Ntub* (**c**; stage ~14; 75% repression, $n = 16$ for 2 ng; 58%, $n = 24$ for 1 ng; 15%, $n = 33$ for 0.5 ng) and *Ngnr1* (**f**; stage ~11; 38% repression, $n = 62$) in *Zic2*-injected embryos. Arrows point to sites of loss. **c** (top), The sensory stripes are most affected as injected Myc–*Zic2* (brown) localizes there most frequently; absence of trigeminal ganglion (bottom). **d**, NCAM expression is unaffected by *Zic2* injection. The injected side is indicated by the brown Myc labelling (stage 14, $n = 7$ unaffected). **g**, Massive ectopic expression of *Ntub* in *Ngnr1*-injected neurulae (stage 13/14; 93% strong ectopic expression, 7% weak expression; $n = 89$). **h**, Co-injected *Zic2* inhibits ectopic neuronal differentiation by *Ngnr1* (7% strong ectopic expression, 60% weak expression, 33% no ectopic expression; $n = 111$). **i**, Ectopic expression of *Ngnr1* (arrows) driven by injected VP16–*Zic2* (stage ~11; 70% ectopic, $n = 30$). **j**, **k**, Ectopic *Ntub*⁺ neuronal differentiation in VP16–*Zic2*-injected (**j**; 100% ectopic, $n = 26$) but not in control sibling (**k**; $n = 20$) tailbud embryos (stage ~24). Ectopic neurons (arrows) are associated with Myc–VP16–*Zic2* protein (brown). All panels show dorsal views except for **c**, bottom, which shows an anterior view, and **j**, **k**, which show lateral views. Percentages in all figures refer to number of embryos showing the specified phenotype. Scale bar, 120 μ m for **a**; 250 μ m for **b**, **d**, **e**, **f**; 180 μ m for **c**; 1 mm for **g**–**i**, and 330 μ m for **j**, **k**.

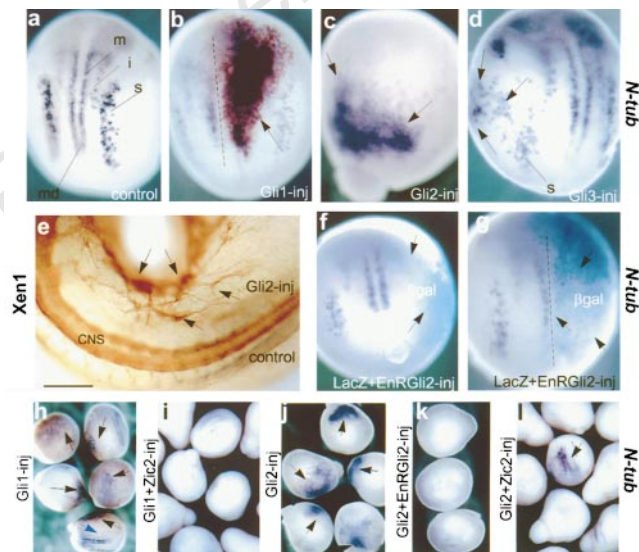


Figure 3 Induction of neurogenesis by Glis and repression by EnR–*Gli2* and *Zic2*. **a**, Normal expression of *Ntub* in primary neuron stripes (stage ~14). **i**, Interneurons; **m**, motor neurons; **md**, midline; **s**, sensory neurons. **b–d**, Ectopic differentiation of *Ntub*⁺ neurons (arrows) in *Gli1*-injected (**b**: 80% ectopic, $n = 15$), *Gli2*-injected (**c**: 95%, $n = 22$ for 2 ng; 87%, $n = 15$ for 1 ng; 74%, $n = 19$ for 0.5 ng; 33%, $n = 21$ for 0.25 ng; 5%, $n = 20$ for 0.125 ng) and *Gli3*-injected embryos (**d**: 63%, $n = 19$ for 2 ng; 32%, $n = 34$ for 1 ng; 11%, $n = 36$ for 0.25 ng). Dashed line in **b**: midline; **s** in **d**: sensory stripe in the injected side. **e**, Ectopic expression (arrows) of *Xen1* in the flank of a tadpole (stage ~34) unilaterally injected with *Gli2*. CNS: central nervous system. **f**, **g**, Repression of endogenous neurogenesis by EnR–*Gli2* in neurulae (stage ~14). The dashed line in **g** marks the midline. β -Galactosidase activity was developed briefly in embryos co-injected with *LacZ* and EnR–*Gli2* RNAs. Thus, the light blue stain pinpoints the injected side but not all the cells that inherited the injected RNAs. **f**, Repression of sensory neurons and interneurons. **g**, Complete repression, including motor neurons, in the injected side. Embryos were scored as positive for repression if one or more of the *Ntub*⁺ stripes were missing (95% suppression in embryos with β -gal in the neural plate, $n = 37$). Embryos with ventral β -gal had normal *Ntub* expression ($n = 3$; not shown). **h**, **j**, Ectopic *Ntub* (arrows) induced by *Gli1* (**h**: stage ~14; 63% ectopic, $n = 62$) and *Gli2* (**j**: stage ~14; 100%, $n = 22$). 5% of injected embryos also showed a repression of the sensory stripe (blue arrow in **h**). **i**, **k**, **l**, Absent or reduced ectopic *Ntub*⁺ neurogenesis (arrows) in embryos co-injected with *Gli1* and *Zic2* (**i**: stage ~14; 13% weak-ectopic expression, 39% normal endogenous pattern; 48% repressed), *Gli2* and EnR–*Gli2* (**k**: stage ~14; 94% normal pattern, 6% weak ectopic) and *Gli2* and *Zic2* (**l**: stage ~14; 23% ectopic expression; 66% very weak ectopic expression; 11% normal endogenous pattern). Affected regions expressed Myc–*Zic2* (**b**, **h**, and not shown). Scale bar, 350 μ m for **a–d**, **f**, **g**; 250 μ m for **e**; and 1.2 mm for **h–l**.

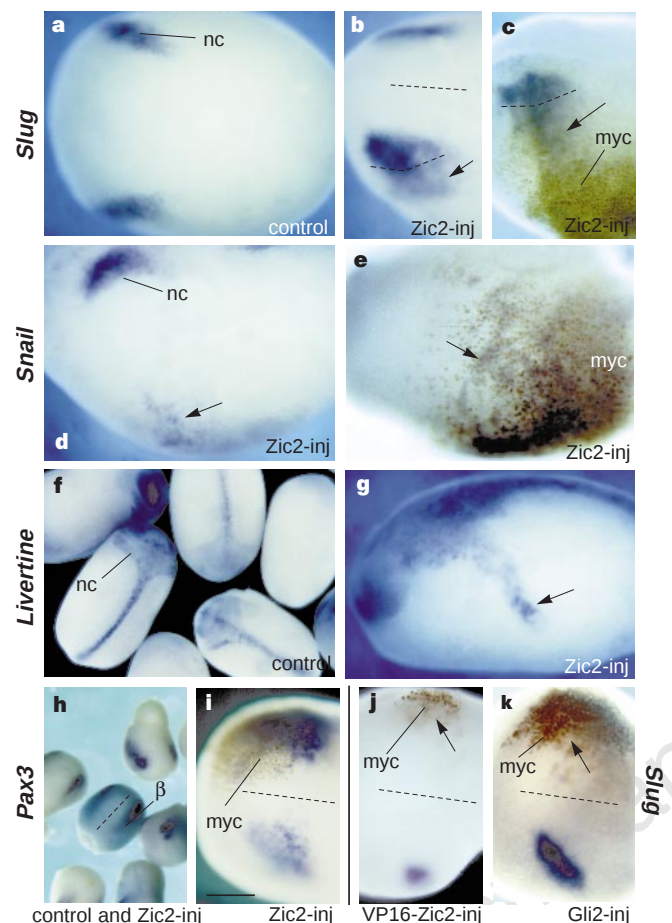


Figure 4 Neural crest induction by Zic2 and repression by VP16-Zic2 and Gli2. **a–g**, Zic2 induces ectopic (arrows) *Slug* (**b, c**: 64% ectopic, $n = 22$), *Snail* (**e**: 43%, $n = 95$ and *Livertine* (**g**: 57%, $n = 30$) not seen in control sibling embryos (**a**: stage ~14; $n = 20$; **d**: stage ~14; $n = 15$; **g**: stage ~24; $n = 52$). **h, i**, Expression of *Pax3* is unchanged in Zic2-injected embryos (stage ~14; $n = 20$). The light blue embryo in **h** was co-injected with Zic2 and LacZ RNAs. Other embryos are uninjected controls. **j**, VP16-Zic2 represses *Slug* (stage ~14, 66% repression, $n = 35$), present in control embryos (not shown; $n = 15$). **k**, Gli proteins inhibit *Slug* (arrows) (stage 13/14, 66% repression, $n = 6$ for Gli2; 33%; $n = 12$ for Gli1 and 60%, $n = 10$ for Gli3; not shown). The injected side was labelled by the expression of Myc-tagged Zic2 protein (myc in **c, e, i–k**) or co-injected β -gal (blue; β in **h**). All panels except **f** and **h** show the anterior end to the left. **a, i–k**, Dorsal view; **b, c, d, g**, lateral view; and **e**, ventral view. The axes of bilateral symmetry are depicted by dashed lines in **b, h–j**, and bilateral expression is shown in **h**. Scale bar, 240 μ m for **a–c, i–k**; 200 μ m for **d, e**; 600 μ m for **f**; 300 μ m for **g**; and 1 mm for **h**.

domains in which the timing of neurogenesis is differentially regulated, and in midline cells, Zic2 may facilitate floorplate differentiation by antagonizing neurogenesis. In the neural folds, Zic2 is a critical regulator of neural crest differentiation, acting downstream of BMP factors³. Its crest-inducing activity could be related to its antineurogenic function, as ectopic expression of the neurogenic bHLH protein Xash3 inhibits crest development⁹. Recruitment of antineurogenic function to the edge of the neural plate could thus underlie two neural innovations with delayed or absent neurogenesis, the forebrain anteriorly and neural crest laterally²³.

Methods

Frog, embryos and microinjection. *Xenopus laevis* albino embryos were injected by standard techniques. Synthetic RNAs (0.2–2 ng per 10 nl per embryo; 2 ng unless otherwise indicated) were microinjected into the animal pole of one cell at the 1–2-cell stage. Synthetic RNAs for microinjection were

made by transcription of pCS2 constructs with SP6 polymerase at 30 °C after digestion with *Not*I in the presence of cap analogue.

cDNA clones. A Zic2 cDNA isolated in our original screen¹² encodes a 501-amino-acid protein (Genbank U57453). Zic2 was subcloned into pBluescript and pCS2-MT in-frame with Myc-epitopes. A PCR fragment of adenovirus VP16 transactivation domain was cloned downstream of the myc epitopes of pCS2-MT2 and upstream of a PCR product encoding the zinc-finger domain of Zic2. fGli1 (ref. 12) was cloned into pCS2-MT. fGli2 was constructed from a full-length clone¹² (clone 9) in pCS2-MT. Our Gli2 clone is allelic to that named Xgli4 (Genbank U42462) (ref. 13). The identity of this clone as the frog homologue of the mouse *Gli2* gene (X99104) is supported by their overall sequence identity (59% in the very amino-terminal region, 92% in the zinc-finger region, and 44% in the very carboxy-terminal region) and their expression pattern¹². Both U42462 and our cDNAs have similar 5' ends but our sequence identifies the initiation methionine at 58 amino acids further upstream, conserving the homology with the mouse gene. The EnR-Gli2 construct was made by cloning a PCR fragment of the five zinc-fingers of Gli2 downstream of the Engrailed repressor domain in pCS2-EnR. fGli3 was constructed in pCS2 from a partial cDNA¹² (clone 10) and an *Apal*–*Kpn*I 3' RACE product derived from Xgli3 (U42461). Our original Gli3 cDNA and U42461 are allelic.

Whole-mount in situ hybridization, immunocytochemistry and histology. RNA probes were made from clones digested and transcribed as follows: pBluescript-Zic2: *Bam*HI and SP6. *Slug*: *Bgl*II and SP6. *Snail*: *Bgl*II and SP6. *Livertine*: *Sall*I and SP6. *N-tubulin*: *Bam*HI and T3. *Pax3*: *Eco*RI and SP6. *Ngnr1*: *Bam*HI and T3, and *Gli3*: *Not*I and T3. Double in situ hybridizations were done with digoxigenin and fluorescein-labelled RNAs followed by Magenta Phos (pink) and nitroblue tetrazolium (NBT) plus 5-bromo-Y-chloro-3-indolyl phosphate (BCIP) (purple) substrates. X-gal reaction (light blue) was performed before in situ hybridization. Myc expression was detected using monoclonal antibody 9E10 (Santa Cruz). Monoclonal antibodies Xen1, Xen2 and Xen3 were used at a dilution of 1:5 (ref. 14). Anti-HNF-3 β polyclonal antibodies² were used at a dilution of 1:8,000. Antibody labelling was detected by peroxidase reactivity with diaminobenzidine (brown/red). Paraplast-embedded embryos were sectioned (14 μ m) in a microtome.

Received 4 February; accepted 2 April 1998.

- Roelink, H. *et al.* Floor plate and motor neuron induction by *vhh-1*, a vertebrate homologue of *hedgehog* expressed by the notochord. *Cell* **76**, 761–775 (1995).
- Ruiz i Altaba, A., Jessell, T. M. & Roelink, H. Restrictions to floor plate induction by *hedgehog* and *winged-helix* genes in the neural tube of frog embryos. *Mol. Cell. Neurosci.* **6**, 106–121 (1995).
- Liem, K. F. Jr, Tremml, G., Roelink, H. & Jessell, T. M. Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. *Cell* **82**, 969–979 (1995).
- Chitnis, A., Henrique, D., Lewis, J., Ish-Horowicz, D. & Kintner, C. Primary neurogenesis in *Xenopus* embryos regulated by a homologue of the *Drosophila* neurogenic gene *Delta*. *Nature* **375**, 761–766 (1995).
- Aruga, J. *et al.* The mouse *Zic* gene family, *J. Biol. Chem.* **271**, 1043–1047 (1996).
- Nagai, T. *et al.* The expression of the mouse *Zic1*, *Zic2* and *Zic3* genes suggests an essential role for *Zic* genes in body pattern formation. *Dev. Biol.* **182**, 299–313 (1997).
- Aruga, J. *et al.* A novel zinc finger protein, *Zic3*, is involved in neurogenesis, especially in the cell lineage of cerebellar granule cells. *J. Neurochem.* **63**, 1880–1890 (1994).
- Ma, Q., Kintner, C. & Anderson, D. J. Identification of *Ngnr1*, a vertebrate neuronal determination gene. *Cell* **87**, 43–52 (1996).
- Ferreiro, B., Kintner, C., Zimmerman, K., Anderson, D. & Harris, W. A. XASH genes promote neurogenesis in *Xenopus* embryos. *Development* **20**, 3649–3655 (1994).
- Han, K. & Manley, J. L. Functional domains of the *Drosophila* engrailed protein. *EMBO J.* **12**, 2723–2733 (1993).
- Kintner, K. W. & Vogelstein, B. The Gli gene encodes a nuclear protein which binds specific sequences in the human genome. *Mol. Cell. Biol.* **10**, 634–642 (1990).
- Lee, J., Platt, K. A., Censullo, P. & Ruiz i Altaba, A. *Gli1* is a target of sonic hedgehog that induces ventral neural tube development. *Development* **124**, 2537–2552 (1997).
- Marine, J.-C., Bellefroid, E. J., Pendevel, H., Martial, J. A. & Pieler, T. A role for *Xenopus* Gli-type zinc finger proteins in the early embryonic patterning of mesoderm and neuroectoderm. *Mech. Dev.* **63**, 211–225 (1997).
- Ruiz i Altaba, A. Cooperation of planar and vertical signals in the induction and patterning of the *Xenopus* nervous system. *Development* **115**, 67–80 (1992).
- Mayor, R., Morgan, R. & Sargent, M. Induction of the prospective neural crest in *Xenopus*. *Development* **121**, 767–777 (1995).
- Essex, L. J., Mayor, R. & Sargent, M. G. Expression of *Xenopus* *Snail* in mesoderm and prospective neural fold ectoderm. *Dev. Dynam.* **198**, 108–122 (1993).
- Ruiz i Altaba, A. & Thery, C. Involvement of *Livertine*, a hepatocyte growth factor family member, in neural morphogenesis. *Mech. Dev.* **60**, 207–220 (1996).
- Mizuseki, K., Kishi, M., Matsui, M., Nakanishi, S. & Sasai, Y. *Xenopus* Zic-related-1 and Sox2, two factors induced by chordin, have distinct activities in the initiation of neural induction. *Development* **125**, 579–587 (1998).
- Nakata, K., Nagai, T., Aruga, J. & Mikoshiba, K. *Xenopus* *Zic3*, a primary regulator both in neural and neural crest development. *Proc. Natl. Acad. Sci. USA* **94**, 11980–11985 (1997).
- Ruiz i Altaba, A. Catching a Gli-mpse of hedgehog. *Cell* **90**, 193–196 (1997).
- Baker, N. E. & Yu, S.-Y. Proneural function of neurogenic genes in the developing *Drosophila* eye. *Curr. Biol.* **7**, 122–132 (1997).

22. Ishibashi, M. *et al.* Targeted disruption of mammalian *hairy* and *Enhancer of split homolog-1 (HES-1)* leads to up-regulation of neural helix-loop-helix factors, premature neurogenesis, and severe neural tube defects. *Genes Dev.* **9**, 3136–3148 (1995).
23. Gans, C. & Northcutt, R. G. Neural crest and the origin of vertebrates: A new head. *Science* **220**, 268–274 (1983).

Acknowledgements. We thank G. Fishell, R. Lehmann, M. Belvin, W. Talbot, N. Dahmane, A. Schier and M. Van Doren for comments; K. Zimmerman for *Pax3*; Q. Ma and D. Anderson for *Ngn1*; D. Turner for the pCS2 series; D. Kessler for pCS2-EnR; D. Melton for NCAM; and J. Aruga and H. Sive for discussion. This work was supported by Skirball Institute start-up funds, a Basil O'Connor Award from the March of Dimes a Pew Fellowship and a grant from the NIH to A.R.A.

Correspondence and requests for materials should be addressed to A.R.A. (e-mail: ria@saturn.med.nyu.edu).

Molecular identification of a hyperpolarization-activated channel in sea urchin sperm

Renate Gauss*, Reinhard Seifert* & U. Benjamin Kaupp

Institut für Biologische Informationsverarbeitung, Forschungszentrum Jülich, D-52425 Jülich, Germany

* These authors contributed equally to this work.

Sea urchin eggs attract sperm through chemotactic peptides, which evoke complex changes in membrane voltage and in the concentrations of cyclic AMP, cyclic GMP and Ca^{2+} ions (see ref. 1 for a review). The intracellular signalling pathways and their cellular targets are largely unknown. We have now cloned, from sea urchin testis, the complementary DNA encoding a channel

polypeptide, SPIH. Functional expression of SPIH gives rise to weakly K^+ -selective hyperpolarization-activated channels, whose activity is enhanced by the direct action of cAMP. Thus, SPIH is under the dual control of voltage and cAMP. The SPIH channel, which is confined to the sperm flagellum, may be involved in the control of flagellar beating. SPIH currents exhibit all the hallmarks of hyperpolarization-activated currents (I_h)^{2,3}, which participate in the rhythmic firing of central neurons, control pacemaking in the heart, and curtail saturation by bright light in retinal photoreceptors^{2,3}. Because of their sequence⁴ and functional properties, I_h channels form a class of their own within the superfamily of voltage-gated and cyclic-nucleotide-gated channels.

We isolated a clone pSPIH from a cDNA library of the testis from sea urchin *Strongylocentrotus purpuratus*. The cDNA encodes a protein of 767 amino-acid residues that shares significant sequence similarity with cyclic-nucleotide-gated (CNG) and voltage-gated K^+ channels (Fig. 1). The SPIH sequence encompasses a voltage sensor (S4) motif, a pore region, and a cyclic-nucleotide-binding domain (S4) motif, a pore region, and a cyclic-nucleotide-binding domain. The S4 segment is characterized by eight positively charged amino acids; the regular spacing of the arginine/lysine residues is interrupted in the midst of the motif (Fig. 1b). The pore loop comprises a GYG motif that is characteristic of pores from K^+ channels^{5,6} (Fig. 1c). However, in contrast to K^+ channels, the SPIH pore motif carries two positively charged residues and a histidine and lacks the cluster of threonine residues that co-determines K^+ selectivity⁶. The cyclic-nucleotide-binding region contains glutamate (E611), arginine (R620) and several glycine residues that have been identified as important determinants of cyclic-nucleotide-binding domains in several proteins⁷ (Fig. 1d).

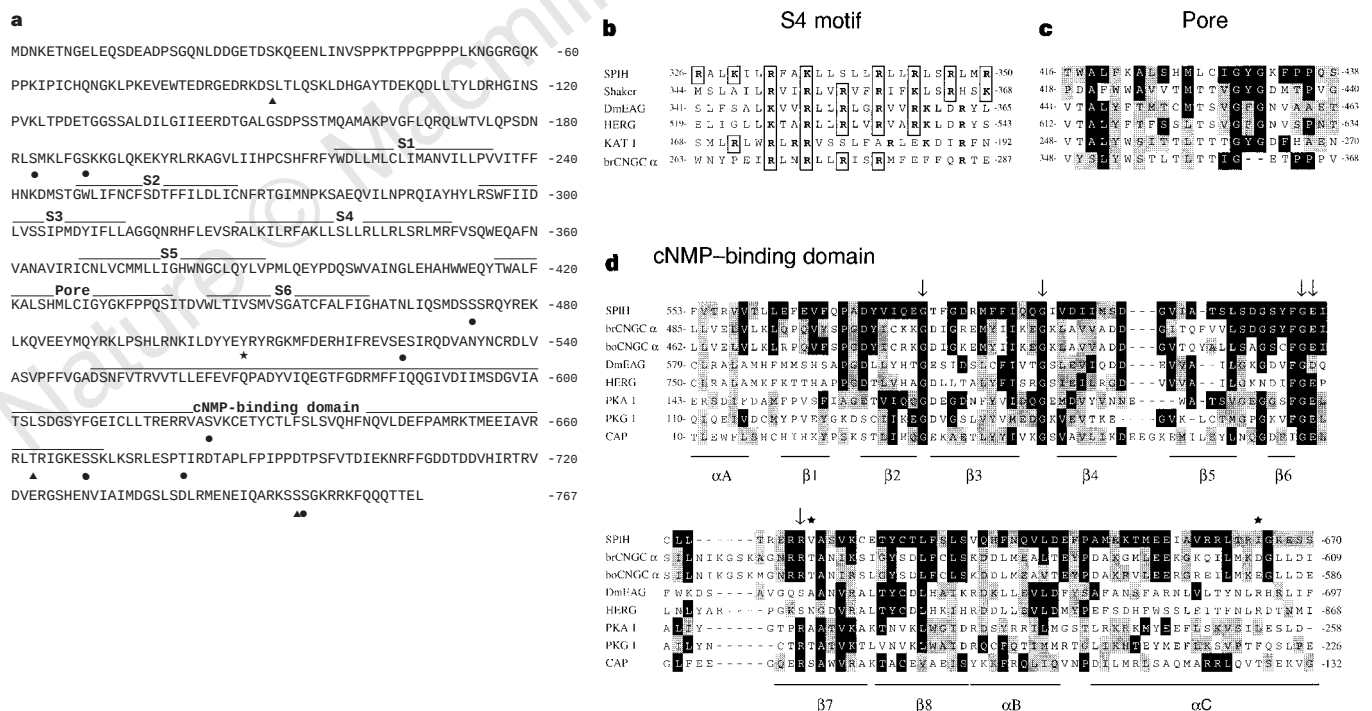


Figure 1 The SPIH channel belongs to the superfamily of voltage- and cyclic-nucleotide-gated channels. **a**, Primary structure of SPIH. The putative six transmembrane segments, S1–S6, the pore region, and the cyclic-nucleotide (cNMP)-binding domain are indicated. Putative sites of phosphorylation by protein kinase A or protein kinase G (triangles), protein kinase C (circles), and tyrosine kinase (asterisk) are indicated. **b**, Comparison of the voltage-sensor (S4) motif of SPIH with that of other channels. Regularly spaced arginine or lysine residues are boxed. Other positively charged residues are in bold. Shaker, K^+ channel encoded by the *Drosophila Shaker b* gene²²; DmEAG, *Drosophila* ether-à-gogo (eag) channel²³; HERG, channel encoded by the human eag-related gene²⁴; KAT1, K^+ channel from *Arabidopsis thaliana*²⁵; brCNGCα, α-subunit of

the CNG channel from bovine rod photoreceptors²⁶. **c**, Comparison of the pore motif of SPIH with that of other channels. Residues identical or similar to the corresponding positions in SPIH are highlighted by a black or grey background, respectively. **d**, Comparison of cNMP-binding domains. brCNGCα, α-subunit of the CNG channel from bovine olfactory neurons²⁷; PKA1, cAMP-binding site 1 of protein kinase A²⁸; PKG1, cGMP-binding site 1 of protein kinase G²⁹; CAP, catabolite gene activator protein³⁰. Residues important for the structure of cNMP-binding domains and for the binding of the ligand are marked by arrows; residues that co-determine the ligand selectivity in brCNGCα are marked by an asterisk. Secondary-structure predictions derived from CAP are shown below the sequences.

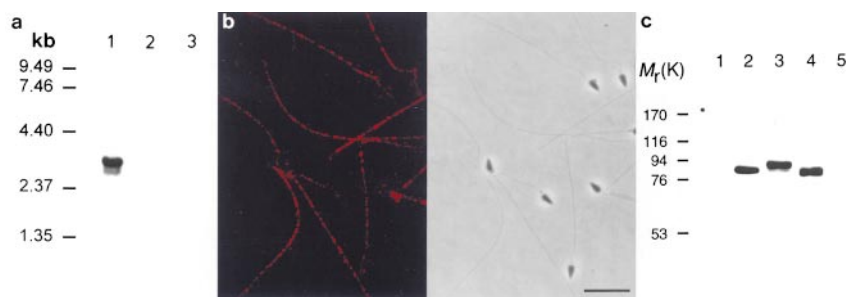


Figure 2 Site of expression of the SPIH channel in *S. purpuratus*. **a**, Northern blot analysis of SPIH messenger RNA in male gonads (lane 1), female gonads (lane 2), and intestine (lane 3); 10 μ g poly(A)⁺ RNA was used in each lane. **b**, Localization of SPIH by immunocytochemistry. Sperm were stained with the antibody FPC44K and with CY-3-conjugated secondary antibody (left). Bar, 10 μ m. **c**, Western blot

analysis of membranes from mock-transfected HEK293 cells (lane 1; 2.5 μ g protein), from HEK293 cells transfected with SPIH cDNA (lane 2; 2.5 μ g protein), from purified flagella of sperm (lane 3; 6 μ g protein), from dephosphorylated flagellar membranes (lane 4; 6 μ g protein), and from sperm heads (lane 5; 15 μ g protein); these membranes were incubated with antibody FPC44K.

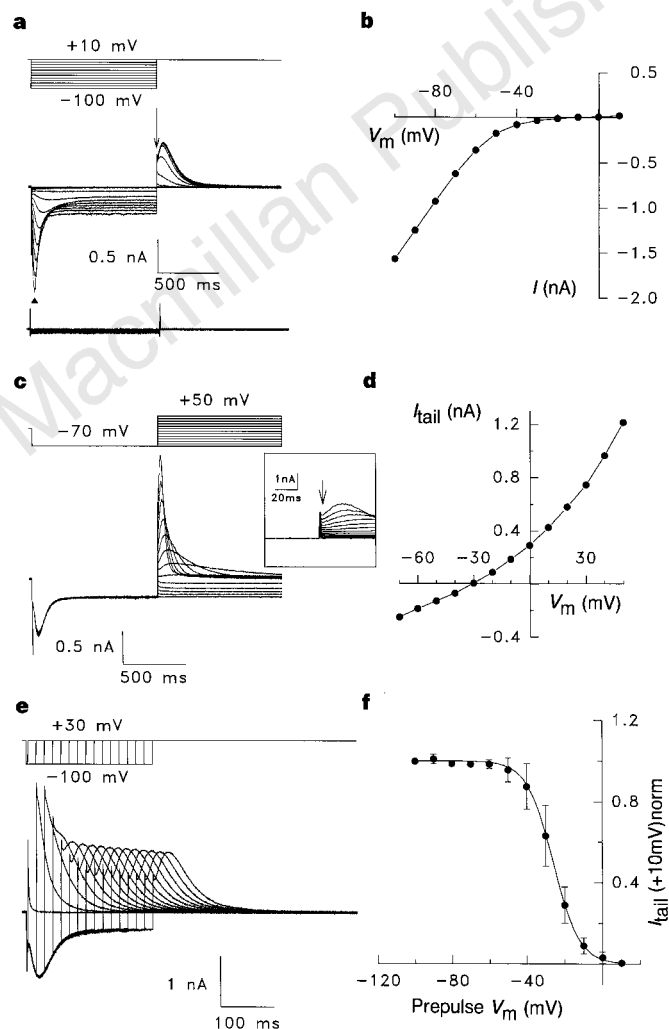


Figure 3 The SPIH channel opens by hyperpolarization. **a**, Whole-cell currents recorded from transfected (upper) and non-transfected (lower) HEK293 cells by stepping the V_m from +10 to -100 mV in increments of 10 mV. Tail currents were recorded by stepping the test voltage back to +10 mV. **b**, Plot of the peak current-voltage (I/V_m) relationship measured at the time indicated by the arrowhead in **a**. **c**, Instantaneous I/V_m protocol; voltage was first stepped from 0 mV to -70 mV. After the current relaxed to a stable plateau, V_m was stepped to test values between -70 mV and +50 mV in 10 mV increments. Inset shows tail currents at an

extended timescale. **d**, Plot of the instantaneous I/V_m relationship measured at the time indicated by the arrow in the inset of **c**. **e**, Waveform and amplitude of tail currents depends on the time at which V_m was stepped back from -100 mV to +30 mV. **f**, Voltage dependence of the relative open probability, P_o . Tail current amplitudes (arrow in part **a**) were normalized to the maximal current. A mean midpoint voltage, $V_{1/2}$, of -26.1 mV and an effective gating charge, Q , of 3.5 elementary charges was determined from a fit of the Boltzmann function to the data; mean of seven experiments.

In northern blot analysis, a major channel transcript of ~ 3.3 kilobases (kb) and a minor transcript of ~ 2.9 kb were detected in male, but not female, gonads. The size of the transcripts concurs with the size of the cloned cDNA (3 kilobase pairs (kbp)). The SPIH-specific probe did not hybridize to poly(A)⁺ RNA from intestine (Fig. 2a). The antibody FPc44K almost exclusively stained the sperm flagellum (Fig. 2b). We confirmed the flagellar localization by western blot analysis of membranes from purified sperm flagella and heads. The antibody FPc44K recognized a band of relative molecular mass $\sim 92,000$ ($M_r \sim 92K$) in flagellar membranes but not in membranes purified from the head of sperm (Fig. 2c, lanes 3, 5). The channel expressed in HEK293 cells has an M_r of $\sim 88K$, which is almost identical to the calculated M_r of 87.9K but

$\sim 4K$ lower than the M_r of the native channel (Fig. 2c, lane 2). Treatment of flagellar membranes with alkaline phosphatase lowered the M_r of the native polypeptide from $\sim 92K$ to $\sim 88K$ (Fig. 2c, lane 4). The 88K band in HEK293 cells was not affected by phosphatase treatment (data not shown). Thus, the native polypeptide exists in a phosphorylated form.

SPIH currents are strikingly similar to hyperpolarization-activated currents observed in many cells^{2,3}. Channels underlying I_h currents become activated by hyperpolarization, are directly modulated by cyclic nucleotides, are blocked by millimolar concentrations of external caesium, are weakly K⁺-selective, and do not carry inward Na⁺ currents in the absence of extracellular K⁺. We examined which of these features are shared by the heterologously expressed SPIH channel. We studied the activity of SPIH channels in transfected HEK293 cells with a patch clamp in the whole-cell configuration. Hyperpolarizing steps in membrane voltage, V_m , from a holding value of +10 mV to more negative test values produced inward currents with a complex waveform (Fig. 3a). First, an instantaneous current was observed. This component was followed by a time-dependent current that developed with a sigmoidal time course. After roughly 20–100 ms, the current reached a peak and then decayed to a much lower plateau. The instantaneous component was often larger than the plateau current (Fig. 3c, e), indicating that it may be, in part, carried by SPIH. Tail currents, observed after stepping V_m back to +10 mV, also exhibited a biphasic waveform. Similar waveforms of I_h tail currents have been recorded from cardiac Purkinje cells⁸.

The peak current/voltage (I/V_m) relationship exhibited a pronounced inward rectification (Fig. 3b). We determined the instantaneous I/V_m relationship (Fig. 3d) from the amplitude of tail currents (Fig. 3c, inset arrow) using a different voltage-step protocol. V_m was first stepped from 0 mV to -70 mV. After a plateau current was reached, V_m was stepped to various test values between -70 mV and +50 mV. The instantaneous I/V_m relationship was outwardly rectifying (Fig. 3d); the extent of the rectification depended on the concentration of extracellular K⁺ (see below). The reversal voltage, V_{rev} , was -30 mV, indicating that SPIH carries a mixed current. The waveform and amplitude of tail currents depended on the time at which V_m was reversed, reflecting time-dependent changes in the open probability, P_0 (Fig. 3e). When V_m was reversed before the inward current peaked, tail currents were large and decayed almost exponentially. When V_m was reversed during the current plateau, tail currents first increased and then decayed to the zero current level. The complex current waveform indicates that the SPIH channel may undergo distinct processes of activation, inactivation, and deactivation. We determined the voltage dependence of relative P_0 (Fig. 3f) from the amplitude of instantaneous tail currents at +10 mV that were like those shown in Fig. 3a (arrow). A fit of the data to the Boltzmann function yielded a mean midpoint voltage at half-maximal current, $V_{1/2}$, of -26.1 mV.

With 1 mM cAMP in the pipette solution, hyperpolarization produced large currents of sigmoidal waveform (Fig. 4a), which is characteristic of the time course of vertebrate I_h currents^{2,3}. cGMP at 1 mM in the pipette enhanced SPIH currents in a similar way (data not shown). We determined the voltage dependence of P_0 in the presence of cAMP from whole-cell tail currents (Fig. 4d, inset, filled circles). A fit of the Boltzmann function to the data yielded a $V_{1/2}$ of -50.8 mV.

To compare SPIH currents in the presence and absence of cyclic nucleotides in the same cell, and to avoid slow effects, such as phosphorylation, we used rapid photorelease of cAMP or cGMP from caged derivatives⁹. Before the flash of ultraviolet light, stepping V_m from +10 mV to -70 mV induced a transient SPIH current (Fig. 4b, upper part, trace 1). After three short flashes of ultraviolet light, the current increased because of the liberation of cAMP, and

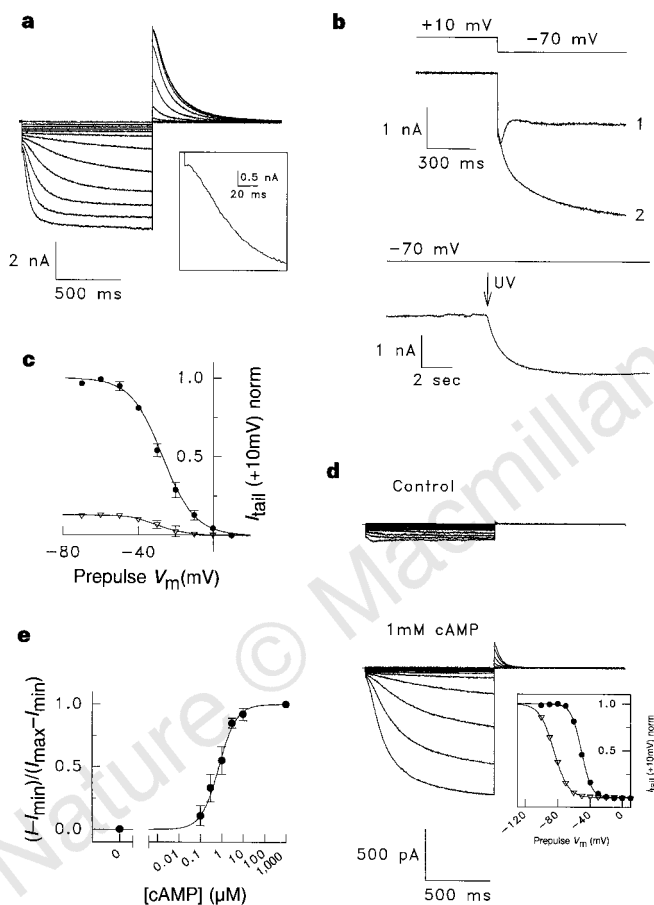


Figure 4 Modulation of the SPIH channel by cAMP. **a**, Whole-cell SPIH currents in the presence of 1 mM cAMP in the pipette. Voltage-step protocol as in Fig. 3a. Inset shows an illustration of the sigmoid time course of current (at -100 mV) at an extended timescale. **b**, Modulation of whole-cell SPIH currents by photolysis of caged cAMP (100 μ M) in the pipette solution. The SPIH current was activated by a voltage step from +10 mV to -70 mV before (trace 1) and after three consecutive ultraviolet (UV) flashes (trace 2). The time course of the flash-induced increase of current at -70 mV is shown below. **c**, Voltage dependence of tail currents normalized to the maximal cAMP-induced current before (triangles) and after (circles) the UV flashes; mean of two experiments. **d**, Voltage-activated SPIH currents in inside-out membrane patches without cAMP (upper) and with 1 mM cAMP (lower) in the bath. Voltage step protocol as in Fig. 3a. Inset shows the voltage dependence of the relative P_0 derived from normalized tail currents at +10 mV, recorded in the whole-cell configuration (from **a**, circles) and from inside-out patches (from **d**, bottom, triangles). Lines represent a fit of the Boltzmann function to the data. $V_{1/2}$ was -50.8 mV (whole-cell configuration) and -84.7 mV (inside-out patch); Q values were 3.8 and 2.7, respectively. **e**, Dependence of the SPIH current on bath cAMP concentration (in μ M: 0.1, 0.3, 1, 3, 10 and 1,000). $V_m = -100$ mV. I_{min} represents the current in the absence of cAMP. Line represents a fit of the Hill equation to the data; $K_{1/2} = 0.74$ μ M; $n = 1.05$; mean of ten experiments.

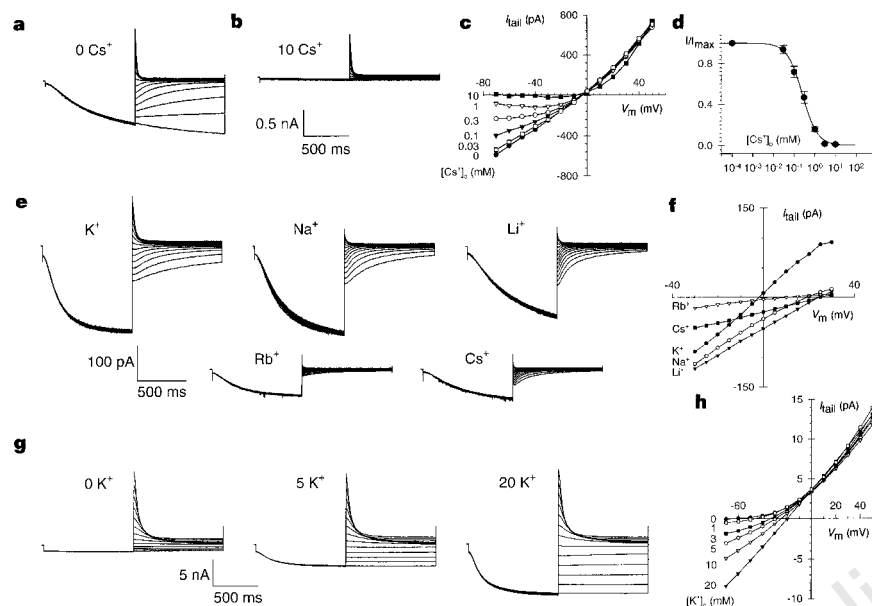


Figure 5 Pharmacological characterization of the SPIH channel. **a, b** Voltage-activated SPIH currents recorded from outside-out membrane patches without (**a**) and with (**b**) 10 mM Cs^+ in the bath, with 1 mM cAMP in the pipette. Voltage-step protocol as in Fig. 3c. **c**, I/V_m relationships, with the indicated $[\text{Cs}^+]_o$ in the bath. **d**, Dependence of normalized current on the extracellular Cs^+ concentration ($[\text{Cs}^+]_o$) at -70 mV. Line represents a fit of the Hill equation to the data; $K_i = 245 \mu\text{M}$, $n = 1.2$; (mean $\pm$ s.d. from one to six experiments). **e**, Ion selectivity of SPIH. **f**, Instantaneous I/V_m relationship from **e**; values for V_{rev} were 16.9 mV (Na^+), 20.6 mV (Li^+), 5.6 mV (Rb^+) and 24.6 mV (Cs^+); means of three to ten experiments. **g**, Whole-cell inward currents in the presence of 0, 5 and 20 mM K^+ in the bath. Ionic strength was adjusted to the same value by the respective concentrations of NMDG-Cl. Voltage-step protocol as in Fig. 3c. **h**, Instantaneous I/V_m relationship with the indicated $[\text{K}^+]_o$ in the bath. $P_{\text{Na}}/P_{\text{K}}$ changed with bath $[\text{K}^+]$: 0.12 (1 mM), 0.2 (3 mM), 0.23 (5 mM), 0.25 (10 mM) and 0.26 (20 mM).

the waveform resembled that of currents recorded with cAMP in the pipette (Fig. 4b, upper part, trace 2). The flash-induced onset of current at constant V_m was rapid (τ of ~ 100 ms; Fig. 4b, lower part), indicating that cAMP acts directly on the channel. Photolysis of caged cGMP (100 μM) did not alter SPIH currents (data not shown). The voltage dependence of normalized tail currents before and after cAMP liberation shows that, without cAMP, the P_0 is $\ll 1$, even at very negative voltages, and that cAMP dramatically enhances P_0 at all voltages (Fig. 4c).

To quantify the cAMP-mediated modulation of P_0 , we recorded SPIH currents from inside-out membrane patches in the presence and absence of cAMP (Fig. 4d). cAMP reversibly increased the voltage-activated currents by up to 20-fold in the absence of Mg^{2+} and ATP. Superfusion of the excised patch with various concentrations of cAMP increased SPIH currents in a dose-dependent fashion, which was described by a simple binding isotherm with a $K_{1/2}$ of 0.74 μM and a Hill coefficient, n , of 1.05 (Fig. 4e). In excised patches, $V_{1/2}$ was significantly more negative than $V_{1/2}$ determined in the whole-cell configuration (Fig. 4d, inset). A similar shift was observed for the cardiac I_h channel¹⁰. $V_{1/2}$ values of vertebrate I_h currents are notoriously variable, and this variability has been attributed to some endogenous factor that might be lost in the excised patch or during dialysis of the cell with the pipette solution^{2,3}. cGMP up to concentrations of 1 mM did not change patch currents. The enhancement of whole-cell currents by cGMP may result from an increase in cAMP levels because of the inhibition of cGMP-regulated phosphodiesterases¹¹.

Extracellular caesium blocked currents in outside-out membrane patches in a concentration- and voltage-dependent fashion. In the presence of 10 mM Cs^+ , inward currents were entirely abolished, whereas outward tail currents were still present (Fig. 5a, b). The I/V_m relationships resulting from use of 0–10 mM Cs^+ are shown in Fig. 5c. Normalized current I/I_{max} is plotted against extracellular Cs^+ concentration ($[\text{Cs}^+]_o$) in Fig. 5d. The Hill equation was fitted to the data with an inhibitory constant, K_i , of 245 μM and a Hill coefficient of $n = 1.2$. Vertebrate I_h currents are blocked by 0.1–5 mM Cs^+ (ref. 3).

We determined the ion selectivity of SPIH in inside-out patches after replacing 100 mM of the bath K^+ (150 mM) with Rb^+ , Na^+ , Li^+ , or Cs^+ (Fig. 5e). We calculated the permeability ratios $P_{\text{K}}:P_{\text{Rb}}:P_{\text{Na}}:P_{\text{Cs}}:P_{\text{Li}}$ to be 1:0.72:0.18:0.06:0.05 from V_{rev} (Fig. 5f). This ion selectivity concurs well with permeability ratios of vertebrate I_h channels^{3,12}. Without K^+ in the extracellular medium,

whole-cell inward but not outward currents were abolished (Fig. 5g). Elevation of extracellular K^+ concentrations dramatically increased inward currents and $P_{\text{Na}}/P_{\text{K}}$; in addition, the instantaneous I/V_m relationship became linear (Fig. 5h). These results indicate that the SPIH channel conducts little, if any, Na^+ in the absence of K^+ ions.

These electrophysiological properties unequivocally identify SPIH as a member of the I_h channel family. However, we also noticed characteristic differences between SPIH and vertebrate I_h channels. First, in the absence of cAMP, the SPIH current is transient, whereas decrements of vertebrate I_h currents during maintained hyperpolarization have not been observed³. Second, the large augmentation of SPIH currents by cAMP arises from an increase in the maximum current, rather than from a positive shift in $V_{1/2}$ as observed for vertebrate I_h channels^{2,3}. Finally, the cardiac I_h is also modulated by cGMP¹³, whereas SPIH is not.

SPIH channels share a characteristic S4 motif with K^+ , Na^+ , and Ca^{2+} channels, which are gated open by depolarization. There is no *a priori* reason why this S4 motif could not be a voltage sensor in a channel activated by hyperpolarization. However, the strong inward rectification of SPIH channels might be produced by a mechanism similar to that of the HERG K^+ channel^{14,15}. HERG channels are closed at rest and open on depolarization. Their strong inward rectification results both from rapid inactivation, which shuts channels off at positive voltages, and from recovery from inactivation at negative voltages. We propose that SPIH is gated by both voltage and cAMP and that the relative P_0 revealed by tail currents reflects steady-state inactivation.

A channel with properties similar to SPIH has been described in *S. purpuratus* sperm¹⁶. The SPIH channel might become activated *in vivo* by speract-induced hyperpolarization¹⁷ and a concomitant rise in cAMP levels¹⁸, and could thereby control the waveform of flagellar beating. □

Methods

Cloning of SPIH. We screened a cDNA library from gonads of the male sea urchin *S. purpuratus* with a DNA fragment that had been amplified by the polymerase chain reaction from cDNA by using degenerate primers. We designed primers specific for sequences that are conserved between cyclic-nucleotide-gated channels and a cDNA clone isolated from the moth *Heliothis*; this clone appeared to encode a member of the CNG-channel family. Several overlapping partial clones were isolated and combined to produce a full-length clone of ~ 3 kbp with an open-reading frame of 767 amino acids.

Northern and western blots. A northern blot was hybridized with a 32 P-labelled 1,074-bp SPIH cDNA fragment. The polyclonal antibody FPC44K, directed against an SPIH fusion protein (amino acids 662–767), was purified from rabbit serum by affinity chromatography. Sperm flagella were separated from the head as described¹⁹. Purified membranes from flagella, heads and HEK293 cells were homogenized in solubilization buffer containing (in mM): 150 NaCl, 1 MgCl₂, 20 HEPES-NaOH at pH 7.5 and 0.1 EGTA, with 0.5% Triton X-100. Proteins were dephosphorylated with 1 unit of alkaline phosphatase for 30 min at 30 °C. Immunoreactivity in western blots was visualized with the electrochemiluminescence detection kit (Amersham). Immunocytochemistry was performed as described²⁰.

Electrophysiology. SPIH was expressed in HEK293 cells as described²¹. Currents were recorded with a patch clamp in the whole-cell configuration and from excised membrane patches^{9,21}. When necessary, leak currents were subtracted off-line using steady-state currents between +10 and +40 mV. Experiments with caged cAMP or caged cGMP were performed as described⁹. In the whole-cell configuration, the bath contained (in mM): 135 NaCl, 5 KCl, 1.8 CaCl₂, 2.8 MgCl₂ and 5 HEPES-NaOH at pH 7.4; the pipette solution contained (in mM): 126 KCl, 10 HEPES-KOH and 10 EGTA at pH 7.4. In the excised-patch configuration, this pipette solution, with the indicated additions, was used on both sides of the patch. Ionic selectivity was determined in inside-out patches by stepping V_m from –70 mV to test values between –30 mV and +30 mV in 5 mV increments. Pipette solution (in mM): 150 KCl, 10 HEPES-NMDG and 10 EGTA at pH 7.4; bath solution (in mM): 50 KCl, 100 XCl, 10 HEPES-NMDG, 10 EGTA at pH 7.4 and 0.1 cAMP. Relative ion permeabilities P_X/P_K were calculated according to the equation $P_X/P_K = \{[K^+]_0 - [K^+]_i \exp(zFV_{rev}/RT)\} / ([X^+]_i \exp(zFV_{rev}/RT))$. V_{rev} was corrected for liquid junction potentials.

Received 24 December 1997; accepted 19 March 1998.

- Ward, C. R. & Kopf, G. S. Molecular events mediating sperm activation. *Dev. Biol.* **158**, 9–34 (1993).
- DiFrancesco, D. Pacemaker mechanisms in cardiac tissue. *Annu. Rev. Physiol.* **55**, 455–472 (1993).
- Pape, H.-C. Queer current and pacemaker: the hyperpolarization-activated cation current in neurons. *Annu. Rev. Physiol.* **58**, 299–327 (1996).
- Santoro, B., Grant, S. G. N., Bartsch, D. & Kandel, E. R. Interactive cloning with the SH3 domain of N-src identifies a new brain specific ion channel protein, with homology to Eag and cyclic nucleotide-gated channels. *Proc. Natl Acad. Sci. USA* **94**, 14815–14820 (1997).
- Heginbotham, L., Abramson, T. & MacKinnon, R. A functional connection between the pores of distantly related ion channels as revealed by mutant K⁺ channels. *Science* **258**, 1152–1155 (1992).
- Heginbotham, L., Lu, Z., Abramson, T. & MacKinnon, R. Mutations in the K⁺ channel signature sequence. *Biophys. J.* **66**, 1061–1067 (1994).
- Finn, J. T., Grunwald, M. E. & Yau, K.-W. Cyclic nucleotide-gated ion channels: an extended family with diverse functions. *Annu. Rev. Physiol.* **58**, 395–426.
- DiFrancesco, D. Characterization of the pace-maker current kinetics in calf Purkinje fibres. *J. Physiol.* **348**, 341–367 (1984).
- Hagen, V. et al. Caged compounds of hydrolysis-resistant analogues of cAMP and cGMP: synthesis and application to cyclic nucleotide-gated channels. *Biochemistry* **35**, 7762–7771 (1996).
- DiFrancesco, D. & Mangoni, M. Modulation of single hyperpolarization-activated channels (I_h) by cAMP in the rabbit sino-atrial node. *J. Physiol.* **474**, 473–482 (1994).
- Beavo, J. A. Cyclic nucleotide phosphodiesterases: functional implications of multiple isoforms. *Physiol. Rev.* **75**, 725–748 (1995).
- Wollmuth, L. P. & Hille, B. Ionic selectivity of I_h channels of rod photoreceptors in tiger salamanders. *J. Gen. Physiol.* **100**, 749–765 (1992).
- DiFrancesco, D. & Tortora, P. Direct activation of cardiac pacemaker channels by intracellular cyclic AMP. *Nature* **351**, 145–147 (1991).
- Trudeau, M. C., Warmke, J. W., Ganetzky, B. & Robertson, G. A. HERG, a human inward rectifier in the voltage-gated potassium channel family. *Science* **269**, 92–95 (1995).
- Smith, P. L., Baukrowitz, T. & Yellen, G. The inward rectification mechanism of the HERG cardiac potassium channel. *Nature* **379**, 833–836 (1996).
- Labarca, P. et al. A cAMP regulated K_v-selective channel from the sea urchin sperm plasma membrane. *Dev. Biol.* **174**, 271–280 (1996).
- Babcock, D. F., Bosma, M. M., Battaglia, D. E. & Darszon, A. Early persistent activation of sperm K⁺ channels by the egg peptide speract. *Proc. Natl Acad. Sci. USA* **86**, 6001–6005 (1992).
- Hansbrough, J. R. & Garbers, D. L. Speract—purification and characterization of a peptide associated with eggs that activates spermatozoa. *J. Biol. Chem.* **256**, 1447–1452 (1981).
- Darszon, A., Labarca, P., Beltrán, C., García-Soto, J. & Liévano, A. Sea urchin sperm: an ion channel reconstitution study case. *Methods: A Companion Methods Enzymol.* **6**, 37–50 (1994).
- Bönigk, W. et al. Rod and cone photoreceptor cells express distinct genes for cGMP-gated channels. *Neuron* **10**, 865–877 (1993).
- Baumann, A., Frings, S., Godde, M., Seifert, R. & Kaupp, U. B. Primary structure and functional expression of a *Drosophila* cyclic nucleotide-gated channel present in eyes and antennae. *EMBO J.* **13**, 5040–5050 (1994).
- Pongs, O. et al. *Shaker* encodes a family of putative potassium channel proteins in the nervous system of *Drosophila*. *EMBO J.* **7**, 1087–1096 (1988).
- Warmke, J., Drysdale, R. & Ganetzky, B. A distinct potassium channel polypeptide encoded by the *Drosophila* eag locus. *Science* **252**, 1560–1562 (1991).
- Warmke, J. W. & Ganetzky, B. A family of potassium channel genes related to eag in *Drosophila* and mammals. *Proc. Natl Acad. Sci. USA* **91**, 3438–3442 (1994).
- Anderson, J. A., Huprikar, S. S., Kochian, L. V., Lucas, W. J. & Gaber, R. F. Functional expression of a probable *Arabidopsis thaliana* potassium channel in *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **89**, 3736–3740 (1992).
- Kaupp, U. B. et al. Primary structure and functional expression from complementary DNA of the rod

- photoreceptor cyclic GMP-gated channel. *Nature* **342**, 762–766 (1989).
- Ludwig, J., Margalit, T., Eismann, E., Lancet, D. & Kaupp, U. B. Primary structure of cAMP-gated channel from bovine olfactory epithelium. *FEBS Lett.* **270**, 24–29 (1990).
- Titani, K. et al. Amino acid sequence of the regulatory subunit of bovine type I adenosine cyclic 3',5'-phosphate dependent protein kinase. *Biochemistry* **23**, 4193–4199 (1984).
- Takio, K. et al. Guanosine cyclic 3',5'-phosphate dependent protein kinase, a chimeric protein homologous with two separate protein families. *Biochemistry* **23**, 4207–4218 (1984).
- Aiba, H., Fujimoto, S. & Ozaki, N. Molecular cloning and nucleotide sequencing of the gene for *E. coli* cAMP receptor protein. *Nucleic Acids Res.* **10**, 1345–1361 (1982).

Acknowledgements. We thank H. Breer for providing a cDNA clone that encodes a *Heliothes* SPIH homologue; J. E. Brown, E. Eismann, and I. Weyand for careful reading of the manuscript; A. Eckert for preparing the manuscript; and H. Schauf for the illustrations. R.G. was the recipient of a stipend from the Boehringer Ingelheim Fonds.

Correspondence and requests for materials should be addressed to U.B.K. (e-mail: a.eckert@fz-juelich.de). The Genbank accession number for the SPIH nucleotide sequence is Y16880.

A family of hyperpolarization-activated mammalian cation channels

Andreas Ludwig, Xiangang Zong, Michael Jeglitsch, Franz Hofmann & Martin Biel

Institut für Pharmakologie und Toxikologie, Technische Universität München, Biedersteiner Strasse 29, 80802 München, Germany

Pacemaker activity of spontaneously active neurons^{1–3} and heart cells^{4–6} is controlled by a depolarizing, mixed Na⁺/K⁺ current, named I_h (or I_f in the sinoatrial node of the heart)^{1,4}. This current is activated on hyperpolarization of the plasma membrane. In addition to depolarizing pacemaker cells, I_h is involved in determining the resting membrane potential of neurons^{1,2} and provides a mechanism to limit hyperpolarizing currents in these cells^{7–9}. Hormones and neurotransmitters that induce a rise in cyclic AMP levels increase I_h by a mechanism that is independent of protein phosphorylation, and which involves direct binding of the cyclic nucleotide to the channel that mediates I_h^{10–13}. Here we report the molecular cloning and functional expression of the gene encoding a hyperpolarization-activated cation channel (HAC1) that is present in brain and heart. This channel exhibits the general properties of I_h channels. We have also identified full-length sequences of two related channels, HAC2 and HAC3, that are specifically expressed in the brain, indicating the existence of a family of hyperpolarization-activated cation channels.

To identify cation channels that are modulated by direct binding of cyclic nucleotides, we performed a BLAST search of the expressed sequence tag (EST) database using the cyclic-nucleotide-binding domain (CNBD) of the cyclic-nucleotide-gated (CNG) channels^{14,15} as a query sequence. We retrieved a sequence (MMA23393) encoding part of a new CNBD and used it to screen a complementary DNA library prepared from mouse brain. We obtained three classes of full-length coding sequences (HAC1–3). The open-reading frames of these sequences predict proteins of 863 (HAC1), 910 (HAC2) and 779 (HAC3) amino acids (Fig. 1a). Primary structure analysis revealed that the proteins have an overall sequence identity of ~60%. HAC1–3 are members of the voltage-gated cation-channel superfamily, members of which are characterized by six membrane-spanning segments (S1–S6), including a voltage-sensing S4 segment, and an ion-conducting pore between S5 and S6. In the carboxy terminus, HAC1–3 contain a sequence exhibiting significant similarity to the CNBD of cyclic-nucleotide-binding proteins such as CNG channels, cAMP-dependent protein kinases and the catabolite-activator protein (CAP) of *Escherichia coli* (Fig. 1b)¹⁴. The pore region of HAC1–3 is related to that of K⁺ channels with a GYG triplet¹⁶. However, the aspartate following the GYG signature sequence in most K⁺ channels is replaced in HAC1–3 by arginine, alanine or glutamine, respectively, indicating that the ion

The voltage-sensing S4 segment of HAC1-3 has a unique structure, consisting of two sequences, each of which contains five positively charged residues; a charged residue is found at every third position. These two sequences are separated by an ‘in-frame’ serine residue. Taken together, HAC1-3 define a new branch of the voltage-gated cation-channel superfamily. HAC1-3 have the highest overall similarity to the CNG channel α -subunits and the *eag* K⁺ channel (28% and 27% sequence identity, respectively; Fig. 1c) of

Using a probe specific for HAC1, we detected a messenger RNA species of 3.4 kilobases in mouse and human brain and heart (Fig. 2a, b). The size of the mRNA concurs with the length of the cloned cDNA. We detected mRNA species of similar size in mouse brain by using HAC2- and HAC3-specific probes (Fig. 2c, d). There was no expression of HAC2 and HAC3 in the heart. The HAC2-specific probe recognized other transcripts of 4.6 kb, 6.2 kb and 8.2 kb. *In situ* hybridization indicated that HAC1 mRNA is highly expressed nearly ubiquitously in brain, with most prominent signals being seen in the hippocampus, thalamus and brain stem (Fig. 2e). Expression of HAC2 is restricted to specific regions of the brain. Strong signals were detected in the hippocampal CA1 region, superior colliculus, cerebral cortex and cerebellum (Fig. 2f). HAC1 mRNA could also be detected in sections of mouse heart ventricles (Fig. 2g).

b

HAC1	FVTAMLLKLFKEVFGPGDYIIREGNIICKMYFIQHGVSIVLTGK---NKE	571
CNG3	LLVELLKLKRFAPVSPGDYICKKGIDLGREMYIIEKGKLAIVAEDE-ITQF	557
PKA	ERLTVDALPEVEQEDHEDKQIVVGCGEEDFFETILLEGSAAVLQRRSENEEF	310
CAP	TLWEFLSHCHIIHKYPKSKSTLLHGGEGKAETLYIVTKSSVAVLIKDE-ECKE	59
	<u>αA</u> <u>β1</u> <u>β2</u> <u>β3</u> <u>β4</u>	
HAC1	MKLS-DG-S-YFGEICLLTR-----GRRTASVRADTYCRIVSLSVDNEN	612
CNG3	VVLG-DG-S-YFGEISILNIKSGSKGNRRNTANIRISIGVSDIFCLSKDGLM	604
PKA	VEVRLGPSDYFGEITALLMN-----RPRAAIVVARGPLKCVKVDPRPFE	354
CAP	MILSYLNQGDFFIGELGLFEEG-----QPSAAIVRAKTACEVAETISYKKER	104
	<u>β5</u> <u>β6</u> <u>β7</u> <u>β8</u> <u>αB</u>	
HAC1	EVLVEEYEMMRRAEETVAIDELDRIGKNSIL	643
CNG3	EALVEYEAKKALEEKGROILMKDNLIDEEL	635
PKA	RVLGPGCSDIILKNIQVNSFVSLSV	379
CAP	QLIQVNPDIILMRSAAMRRLQVTSEKVGNI	135
	<u>αC</u>	

c

```
graph LR;
    eag --- Node1;
    Node1 --- CNG3;
    Node1 --- Node2;
    Node2 --- HAC3;
    Node2 --- Node3;
    Node3 --- HAC2;
    Node3 --- HAC1;
```

588

single exponential with a time constant of 241 ± 12 ms at -140 mV ($n = 11$). The activation of HAC1 revealed a steep voltage dependence, reaching a half-maximal value ($V_{1/2}$) at about -100 mV (Fig. 3d).

We determined the ion selectivity of HAC1 by measuring the current/voltage (I/V) relationship of the fully activated channel (Fig. 3e–h). We determined reversal potentials (E_{rev}) of -32 ± 2 mV ($n = 7$) and -19 ± 1 mV ($n = 14$) at 5.4 mM and 30 mM extracellular K^+ , respectively. These reversal potentials were significantly more positive than would be predicted for a pure potassium conductance (-81 mV and -37 mV at 5.4 mM and 30 mM extracellular K^+ , respectively) indicating that HAC1, like the native I_h

channel^{4,18,19}, passes both Na^+ and K^+ . The relative permeability ratio for Na^+ and K^+ (P_{Na}/P_K), as determined by the Goldman–Hodgkin–Katz equation²⁰, was 0.24 and 0.31 at 5.4 mM and 30 mM extracellular K^+ , respectively. Thus, the kinetic and voltage dependence of current activation and the ion selectivity of HAC1 are compatible with the properties of I_h in a variety of neurons^{1,3,13} and heart cells^{4,19,21}.

We next investigated whether the HAC1 current, like native I_h , is enhanced by direct interaction with cAMP and cGMP^{10,12,22}. Both cyclic nucleotides increased the whole-cell current amplitude by shifting the activation curve by up to 12 – 14 mV towards positive voltages (Fig. 4a). The $V_{1/2}$ values were -103 ± 1 mV ($n = 11$) in

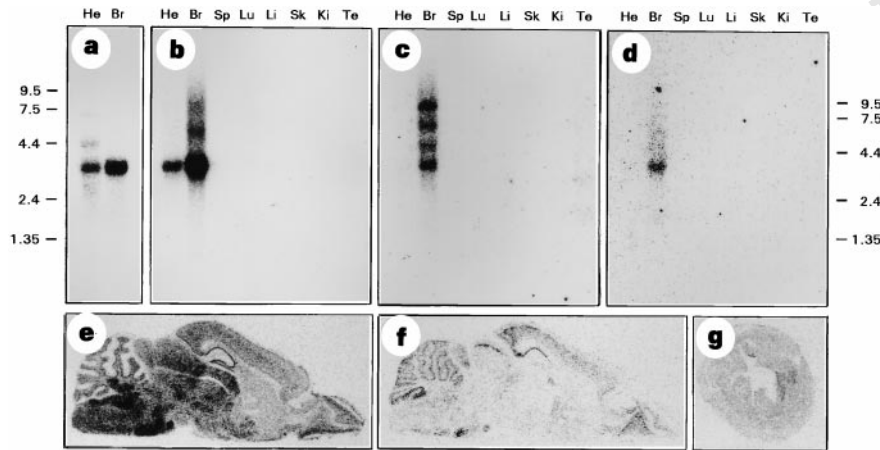


Figure 2 Expression of HAC1–3 mRNA. **a**, Northern blot of human heart (He) and brain (Br) mRNA labelled with an HAC1-specific probe. **b–d**, Northern blot analysis of mRNA from mouse heart (He), brain (Br), spleen (Sp), lung (Lu), liver (Li), skeletal muscle (Sk), kidney (Ki) and testis (Te). The blot was labelled with a probe directed against HAC1 (**b**), HAC2 (**c**), or HAC3 (**d**). Exposure time on

phosphorimager plates was 6 h for **a–c**, and 36 h for **d**. Molecular mass standards (in kb) are indicated on the left for **a** and on the right for **b–d**. **e, f**, *In situ* hybridization analysis of mouse brain sagittal sections labelled with riboprobes directed against HAC1 (**e**) and HAC2 (**f**). **g**, Frontal section through mouse heart ventricles hybridized with a HAC1 riboprobe.

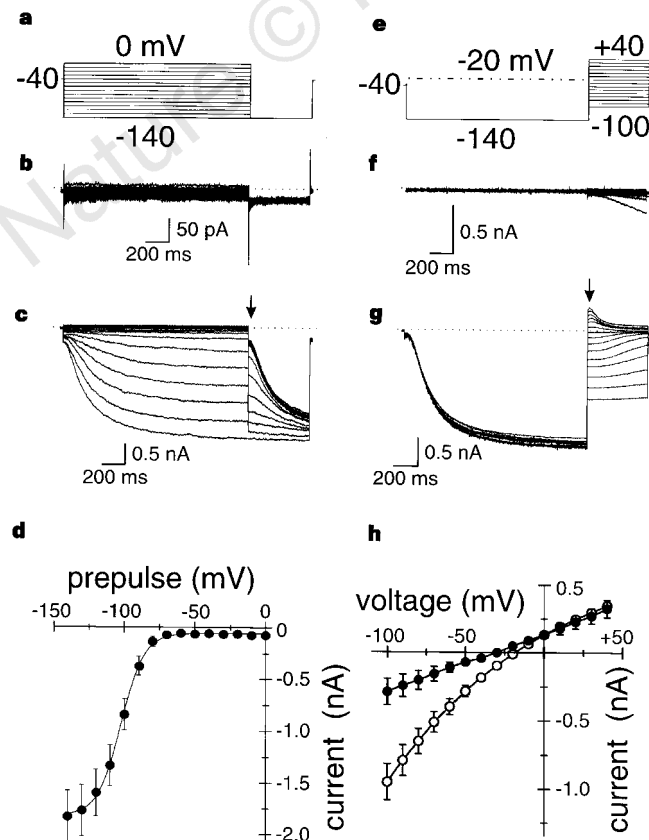


Figure 3 Electrophysiological properties of the expressed HAC1 channel measured in whole-cell voltage clamp. **a–d**, Determination of the voltage dependence of HAC1 activation. **a**, Cells were voltage-clamped from a holding potential of -40 mV to the various voltages (ranging from -140 mV to 0 mV, in 10 -mV increments); this was followed by a step to -140 mV. **b, c**, Current traces measured in a HEK293 cell transfected with empty expression vector (**b**) and the HAC1-expression vector (**c**). **d**, Tail current analysis of HAC1 activation. The amplitude of tail-current traces measured immediately after the voltage step to -140 mV (indicated by the arrow in Fig. 3c) was plotted as function of the preceding membrane potential. The data were fitted by the Boltzmann equation with $A1 = -0.06$ nA, $A2 = -1.82$ nA, $V_{1/2} = -103$ mV and $k = -0.12$ (see Methods). The leak current was digitally subtracted. **e–g**, Determination of the I/V relationship for fully activated HAC1. **e**, Voltage protocol. Steps to test voltages ranging from -100 mV to $+40$ mV, in 10 -mV increments, were either delivered after HAC1 was fully activated for 1.5 s by a step from -40 mV to -140 mV (protocol A) or after a step from -40 mV to -20 mV (protocol B). **f**, Current traces obtained using protocol B. **g**, Current traces obtained using protocol A. **h**, I/V relationship for the fully activated channel obtained by subtracting the tail-current amplitudes of the two sets of tail currents. The tail current amplitudes were measured immediately after the voltage clamp to the indicated test voltages was settled (arrow in Fig. 3g). I/V relationships were determined at 5.4 mM (filled circles) and 30 mM (open circles) extracellular K^+ , respectively.

the absence of cyclic nucleotides, and -90 ± 3 mV ($n = 10$) and -89 ± 3 mV ($n = 7$) in the presence of 1 mM cAMP and 1 mM cGMP, respectively. The steepness of the activation curve was not significantly affected by cAMP and cGMP. cAMP and cGMP did not further increase the current when it was fully activated by hyperpolarization. To determine the cAMP sensitivity of HAC1, we measured cAMP-induced shifts of the activation curve in excised inside-out patches for different cAMP concentrations (Fig. 4b). Fitting the data to a Hill equation yielded values of $0.5 \mu\text{M}$ for the half-maximal cAMP concentration (K_a) and 0.8 for the Hill coefficient

(ν). Both values are in good agreement with properties of native I_h (ref. 10).

cAMP altered not only the voltage dependence of current activation, but also significantly accelerated the kinetics of channel opening (Fig. 4c inset). The potentiation of the current by cAMP was readily reversible by washing out the cyclic nucleotide from the patch, and persisted in the presence of the nonspecific protein kinase inhibitor H-89 (Fig. 4c). Thus cAMP affects HAC1 by directly interacting with the CNBD of the channel protein. Like cAMP, cGMP also enhanced both current amplitude and rate of activation in excised inside-out patches (Fig. 4c). The apparent affinity and the Hill coefficient for cGMP were $K_a = 6 \mu\text{M}$ and $\nu = 1.2$ ($n = 8$), respectively.

As expected for a channel that mediates I_h (refs 1,4,13,21), the current measured in whole-cell voltage clamp was almost completely abolished by 2 mM extracellular Cs^+ , but was not sensitive to 20 mM extracellular tetraethylammonium (TEA) or 1 mM 4-aminopyridine. Similarly, 2 mM Ba^{2+} only slightly reduced the HAC1 current (Fig. 4d, e).

Our results indicate that HAC1 is the first member of a class of channels that are dually gated by hyperpolarization of the membrane and by direct binding of cyclic nucleotides. The functional properties of the HAC1 current, that is, the voltage-dependence of activation, ion selectivity, pharmacological profile and modulation by cyclic nucleotides, concur with the general criteria that characterize I_h in several neuronal and non-neuronal cells^{1,4}. The expression pattern of HAC1 indicates that it may mediate the current that is involved in control of pacemaker activity in both central nervous system and cardiac cells. The identification of two other, brain-specific, members of the HAC family, HAC2 and HAC3, is consistent with the diversity of I_h currents detected in different types of neurons¹. □

Methods

Molecular cloning of HAC1-3. A 217-base-pair fragment (nucleotides 11–228 of the EST MMA23393) was amplified by reverse transcription with polymerase chain reaction from mouse brain mRNA. We used the ^{32}P -labelled fragment to screen 1×10^6 colonies of an oligo(dT)-primed cDNA library from mouse brain. The library was constructed using pcDNAII vector (Invitrogen). From a total of 120 positive clones, 36 were randomly chosen and analysed by restriction mapping and partial sequencing. The clones fell into three separate classes, designated HAC1–3. We sequenced both strands of the two longest clones from each class.

Northern blot. Northern blots (Clontech) containing $2 \mu\text{g}$ mRNA were hybridized under high stringency with ^{32}P -labelled probes. A human multiple tissue blot was hybridized with a fragment corresponding to amino acids 575–647 of HAC1. The mouse tissue blot was hybridized sequentially with probes corresponding to amino acids 98–209 of HAC1, 636–722 of HAC2 and 131–197 of HAC3.

In situ hybridization. *In situ* hybridization was done as described²³. Cryostat sections from BALB/c mice brains and hearts were hybridized with riboprobes labelled with [^{35}S]UTP or [^{33}P]UTP, respectively. Slides were exposed to film for four days. RNA probes were transcribed from fragments corresponding to amino acids 98–209 (5' probe) and 653–741 (3' probe) of HAC1, and amino acids 7–101 (5' probe) and 636–722 (3' probe) of HAC2. Both 5' probes (Fig. 1e, f) and 3' probes of each HAC subtype gave identical results.

Expression and electrophysiological analysis of HAC1. The complete coding region of HAC1 was cloned into the pcDNA3 expression vector (Invitrogen). Transfection of HEK293 cells was described²⁴. Currents were measured 2–4 days after transfection, either in whole-cell mode or in excised inside-out patches at room temperature. For all measurements, except in part of the experiment shown in Fig. 3h, the following solutions were used: extracellular solution (bath in whole-cell mode, pipette in inside-out mode), 110 mM NaCl, 0.5 mM MgCl_2 , 1.8 mM CaCl_2 , 5 mM HEPES, 30 mM KCl, adjusted to pH 7.4 with NaOH; intracellular solution (pipette in whole-cell mode, bath in inside-out mode), 130 mM KCl, 10 mM NaCl, 0.5 mM MgCl_2 , 1 mM EGTA, 5 mM HEPES, adjusted to pH 7.4 with KOH. For the determination

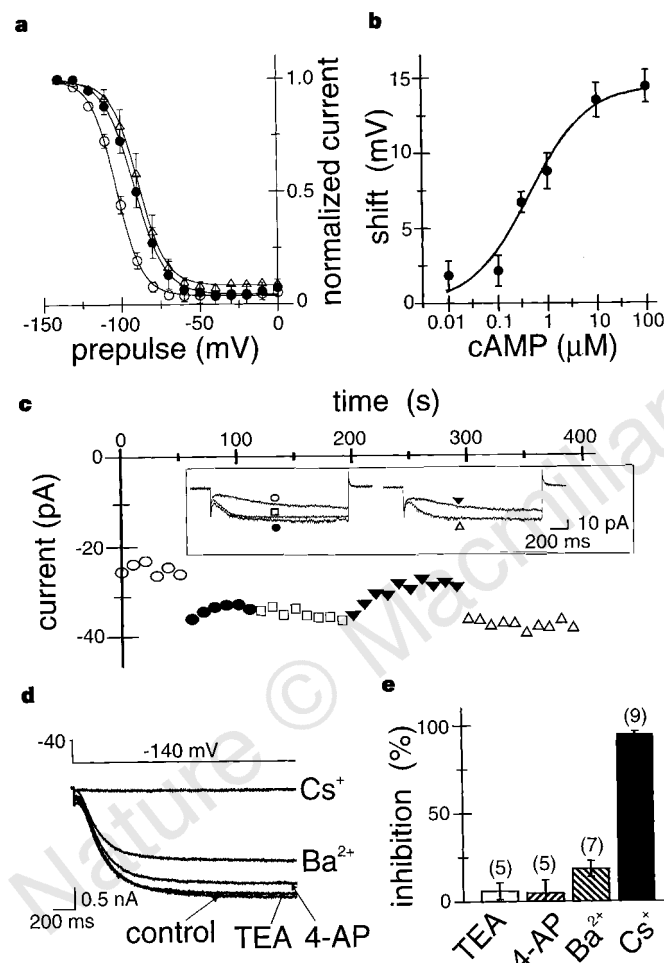


Figure 4 Modulation of the HAC1 current by cyclic nucleotides and channel blockers. **a**, Voltage dependence of activation measured with tail currents in whole-cell mode using the same protocol as in Fig. 3a–d. Currents were determined in the absence of cyclic nucleotides (open circles) or after intracellular perfusion for 1 min with 1 mM cAMP (filled circles) or 1 mM cGMP (triangles), respectively. The curves represent best fits of tail currents using Boltzmann functions. **b**, Dose–response relationship for the shift in the HAC1 activation curve as a function of cAMP concentration. Shifts were determined in inside-out patches ($n = 10$); see Methods. **c**, Increase in the steady-state current by cAMP and cGMP. An excised inside-out patch from a HAC1-expressing cell was perfused with bath solution containing no cyclic nucleotide (open ovals), $10 \mu\text{M}$ cAMP (filled ovals), $10 \mu\text{M}$ cAMP + $1 \mu\text{M}$ protein kinase inhibitor H-89 (squares), no cyclic nucleotide (filled triangles) and $100 \mu\text{M}$ cGMP (open triangles). Currents were evoked by stepping from a holding potential of -40 mV to -130 mV and plotted against time. The inset shows examples of current traces measured under various conditions. **d**, Extracellular blockage of whole-cell current by 2 mM Cs^+ , 2 mM Ba^{2+} , 20 mM tetraethylammonium (TEA) and 1 mM 4-aminopyridine (4-AP), respectively. **e**, Percentage inhibition of steady-state current at -140 mV by 20 mM TEA, 1 mM 4-AP, 2 mM Ba^{2+} , and 2 mM Cs^+ .

of the relative Na^+/K^+ permeability (Fig. 3h), we used, in addition to the standard solution, an extracellular solution with a low K^+ concentration containing 5.4 mM KCl and 135 mM NaCl. Data were acquired at 1 kHz using a List EPC7 amplifier and pCLAMP software (Axon). Average values are mean $\pm$ s.e.m.

Determination of cAMP sensitivity of HAC1. The sensitivity of HAC1 to cAMP was investigated in excised inside-out patches by determining the dose-response relationship for the shift of the activation curve as a function of the cAMP concentration. I/V relationships for HAC1 activation were determined with cAMP concentrations in the bath solution ranging from 0.01 μM to 100 μM . The voltage protocol described in Fig. 3a was used, the only difference being that the hyperpolarizing voltage steps were extended to -150 mV and that the tail currents were measured after stepping to -150 mV. This extension to more negative voltages was needed to fully activate the channel, as in inside-out patches $V_{1/2}$ was shifted by about 20 mV to more hyperpolarizing voltages with respect to the $V_{1/2}$ measured in whole-cell mode. The shift in $V_{1/2}$ did not affect the kinetics of current activation or the steepness of voltage dependence. A similar phenomenon has been described for the wild-type I_h current of sinoatrial node cells²⁵ and may be due to channel run-down or the loss of an intracellular factor that regulates voltage dependence of activation. Amplitudes of tail currents obtained immediately after stepping from the hyperpolarizing prepulses to -150 mV were fitted using the Boltzmann equation $y = [(A1 - A2)/(1 + \exp(V - V_{1/2})/k)] + A2$. $V_{1/2}$ was obtained for each cAMP concentration from the Boltzmann fit. Finally, the shift in $V_{1/2}$ with respect to the $V_{1/2}$ measured in the absence of cAMP was plotted against the respective cAMP concentration. □

Received 4 March; accepted 31 March 1998.

1. Pape, H.-C. Queer current and pacemaker: the hyperpolarization-activated cation current in neurons. *Annu. Rev. Physiol.* **58**, 299–327 (1996).
2. Halliwell, J. V. & Adams, P. R. Voltage-clamp analysis of muscarinic excitation in hippocampal neurons. *Brain Res.* **250**, 71–92 (1982).
3. McCormick, D. A. & Pape, H.-C. Properties of a hyperpolarization-activated cation current and its role in rhythmic oscillation in thalamic relay neurons. *J. Physiol.* **431**, 291–318 (1990).
4. DiFrancesco, D. Pacemaker mechanisms in cardiac tissue. *Annu. Rev. Physiol.* **55**, 455–472 (1993).
5. Brown, H. F., DiFrancesco, D. & Noble, S. J. How does adrenaline accelerate the heart? *Nature* **280**, 235–236 (1979).
6. DiFrancesco, D. A new interpretation of the pace-maker current in calf Purkinje fibres. *J. Physiol.* **314**, 359–376 (1981).
7. Solomon, J. S. & Nerbonne, J. M. Hyperpolarization-activated currents in isolated superior colliculus-projecting neurons from rat visual cortex. *J. Physiol.* **462**, 393–420 (1993).
8. Mayer, M. L. & Westbrook, G. L. A voltage-clamp analysis of inward (anomalous) rectification in mouse spinal sensory ganglion neurons. *J. Physiol.* **340**, 19–45 (1983).
9. Bayliss, D. A., Viana, F., Bellingham, M. C. & Berger, A. J. Characteristics and postnatal development of a hyperpolarization-activated inward current in rat hypoglossal motoneurons *in vitro*. *J. Neurophysiol.* **71**, 119–128 (1994).
10. DiFrancesco, D. & Tortora, P. Direct activation of cardiac pacemaker channels by intracellular AMP. *Nature* **351**, 145–147 (1991).
11. Pape, H.-C. & McCormick, D. A. Noradrenalin and serotonin selectively modulate thalamic burst firing by enhancing a hyperpolarization-activated cation current. *Nature* **340**, 715–718 (1989).
12. McCormick, D. A. & Pape, H.-C. Noradrenergic and serotonergic modulation of a hyperpolarization-activated cation current in thalamic relay neurons. *J. Physiol.* **431**, 319–342 (1990).
13. Pedarzani, P. & Storm, J. F. Protein kinase A-independent modulation of ion channels in the brain by cyclic AMP. *Proc. Natl Acad. Sci. USA* **92**, 11716–11720 (1995).
14. Zagotta, W. N. & Siegelbaum, S. A. Structure and function of cyclic nucleotide-gated channels. *Annu. Rev. Neurosci.* **19**, 235–263 (1996).
15. Biel, M. *et al.* Another member of the cyclic nucleotide-gated channels family, expressed in testis, kidney, and heart. *Proc. Natl Acad. Sci. USA* **91**, 3505–3509 (1994).
16. Heginbotham, L., Lu, Z., Abramson, T. & MacKinnon, R. Mutations in the K^+ channel signature sequence. *Biophys. J.* **66**, 1061–1067 (1994).
17. Santoro, B., Grant, S. G. N., Bartsch, D. & Kandel, E. R. Interactive cloning with the SH3 domain of N-src identifies a new brain specific ion channel protein, with homology to Eag and cyclic nucleotide-gated channels. *Proc. Natl Acad. Sci. USA* **94**, 14815–14820 (1997).
18. Ho, W.-K., Brown, H. F. & Noble, D. High selectivity of the i_f channel to Na^+ and K^+ in rabbit isolated sinoatrial node cells. *Pflügers Arch.* **426**, 68–74 (1994).
19. Yu, H., Chang, F. & Cohen, I. S. Pacemaker current i_f in adult canine cardiac ventricular myocytes. *J. Physiol.* **485**, 469–483 (1995).
20. Hille, B. *Ionic Channels of Excitable Membranes* (Sinauer Associates, Sunderland, MA, 1992).
21. Baker, K., Warren, K. S., Yellen, G. & Fishman, M. C. Defective “pacemaker” current (i_h) in a zebrafish mutant with a slow heart rate. *Proc. Natl Acad. Sci. USA* **94**, 4554–4559 (1997).
22. Pape, H.-C. & Mager, R. Nitric oxide controls oscillatory activity in thalamocortical neurons. *Neuron* **9**, 441–448 (1992).
23. Ludwig, A., Flockerzi, V. & Hofmann, F. Regional expression and cellular localization of the α_1 and β subunit of high voltage-activated calcium channels in rat brain. *J. Neurosci.* **17**, 1339–1349 (1997).
24. Biel, M., Zong, X., Ludwig, A., Sautter, A. & Hofmann, F. Molecular cloning and expression of a modulatory subunit of the cyclic nucleotide-gated cation channel. *J. Biol. Chem.* **271**, 6349–6355 (1996).
25. DiFrancesco, D. & Mangoni, M. Modulation of single hyperpolarization-activated channels (i_f) by cAMP in the rabbit sino-atrial node. *J. Physiol.* **474**, 473–482 (1994).
26. Titani, K. *et al.* Amino acid sequence of the regulatory subunit of bovine type I adenosine cyclic 3',5'-phosphate dependent protein kinase. *Biochemistry* **23**, 4193–4199 (1984).
27. Weber, I. T. & Steitz, T. A. Structure of a complex of catabolite gene activator protein and cyclic AMP refined at 2.5 Å. *J. Mol. Biol.* **198**, 311–326 (1987).

28. Warmke, J. W. & Ganetzký, B. A family of potassium channel genes related to *eag* in *Drosophila* and mammals. *Proc. Natl Acad. Sci. USA* **91**, 3438–3442 (1994).

Acknowledgements. We thank P. Mayr, B. Lehnert, S. Stief and S. Ehrhard for technical support. Research was supported by grants from Deutsche Forschungsgemeinschaft and Fond der Chemischen Industrie.

Correspondence and requests for material should be addressed to M.B. (e-mail: biel@ipt.med.tu-muenchen.de). The Genbank accession numbers for HAC cDNA sequences are AJ225122 (for HAC1), AJ225123 (for HAC2) and AJ225124 (for HAC3).

The chemokine receptor CXCR4 is essential for vascularization of the gastrointestinal tract

Kazunobu Tachibana*, Seiichi Hirota†‡, Hisashi Iizasa*‡§, Hisahiro Yoshida‡||, Kenji Kawabata*, Yuki Kataoka*, Yukihiro Kitamura†, Kouji Matsushima§, Nobuaki Yoshida*, Shin-ichi Nishikawa||, Tadimitsu Kishimoto† & Takashi Nagasawa*

* Department of Immunology, Research Institute, Osaka Medical Center for Maternal and Child Health, 840 Murodo-cho, Izumi, Osaka 590-02, Japan

† Department of Pathology, Osaka University Medical School, 2-2 Yamada-oka, Suita, Osaka 565, Japan

§ Department of Molecular Preventive Medicine, Division of Social Medicine, Graduate School of Medicine, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113, Japan

|| Department of Molecular Genetics, Faculty of Medicine, Kyoto University, Yoshida, Sakyo-ku, Kyoto 606, Japan

¶ Department of Medicine III, Osaka University Medical School, 2-2 Yamada-oka, Suita-city, Osaka 565, Japan

‡ These authors contributed equally to this work.

Vascularization of organs generally occurs by remodelling of the preexisting vascular system during their differentiation and growth to enable them to perform their specific functions during development. The molecules required by early vascular systems, many of which are receptor tyrosine kinases and their ligands, have been defined by analysis of mutant mice^{1–3}. As most of these mice die during early gestation before many of their organs have developed, the molecules responsible for vascularization during organogenesis have not been identified. The cell-surface receptor CXCR4 (refs 4–6) is a seven-transmembrane-spanning, G-protein-coupled receptor for the CXC chemokine PBSF/SDF-1 (for pre-B-cell growth-stimulating factor/stromal-cell-derived factor), which is responsible for B-cell lymphopoiesis, bone-marrow myelopoiesis and cardiac ventricular septum formation⁷. CXCR4 also functions as a co-receptor for T-cell-line tropic human immunodeficiency virus HIV-1 (ref. 8). Here we report that CXCR4 is expressed in developing vascular endothelial cells, and that mice lacking CXCR4 or PBSF/SDF-1 have defective formation of the large vessels supplying the gastrointestinal tract. In addition, mice lacking CXCR4 die *in utero* and are defective in vascular development, haematopoiesis and cardiogenesis, like mice lacking PBSF/SDF-1, indicating that CXCR4 is a primary physiological receptor for PBSF/SDF-1. We conclude that PBSF/SDF-1 and CXCR4 define a new signalling system for organ vascularization.

CXCR4 was identified by screening chemokine receptor-like ‘orphans’ for their ability to respond to PBSF/SDF-1 (refs 4–6), but it remains unresolved whether CXCR4 is the only receptor for PBSF/SDF-1. Moreover, most chemokine receptors bind more than one ligand⁹, so there may be additional ligands for CXCR4.

To elucidate the physiological functions of CXCR4, we generated mice lacking CXCR4 expression. We made a vector to target the

CXCR4 gene by deleting most of exon 2, including all of the transmembrane regions (Fig. 1). Mice with a heterozygous CXCR4 mutation appeared to be healthy and fertile. Homozygous mutant (CXCR4^{-/-}) embryos were present at the expected ratios until day 15.5 of embryogenesis (E15.5). However, about half the CXCR4^{-/-} embryos were dead at E18.5 and CXCR4^{-/-} neonates died within an hour, as noted previously in mice lacking PBSF/SDF-1 (ref. 7).

To study the functions of CXCR4 during embryogenesis, we examined its expression in developing embryos by *in situ* hybridization. High levels of CXCR4 transcripts were detected in the endothelium of developing blood vessels during embryonic development (see below). We next investigated the effect of CXCR4 gene ablation on vascular formation. On visual inspection, vitellin and umbilical vessels appeared to be normal (results not shown). Histological examination of E18.5 CXCR4^{-/-} embryos demonstrated the presence of the major blood vessels, including aorta, vena cava, carotid artery, jugular vein, coeliac artery and superior mesenteric artery (SMA) and vein (SMV).

To visualize the vascularization of organs, whole-mount preparations of wild-type and mutant embryos were immunostained with a monoclonal antibody against the endothelial marker PECAM-1 (refs 10, 11). In the gastrointestinal tract (stomach, intestine and mesentery), a highly branched homogeneous vascular network was observed in both wild-type and mutant embryos by E11.5. Remodelling into large and small vessels in the mesentery connecting mid-gut loop occurred around E12.5 for wild-type embryos and E13.5 wild-type embryos had many large branches of SMA or SMV to supply the intestines (Fig. 2a). However, the mesenteries of all E13.5 *CXCR4*^{-/-} embryos ($n = 8$) contained smaller branches instead of these large branches (Fig. 2b). Mesenteries of E13.5 wild-type embryos had the SMA, SMV and their paired branches; each branch contained the artery and the vein (Fig. 2c). In contrast,

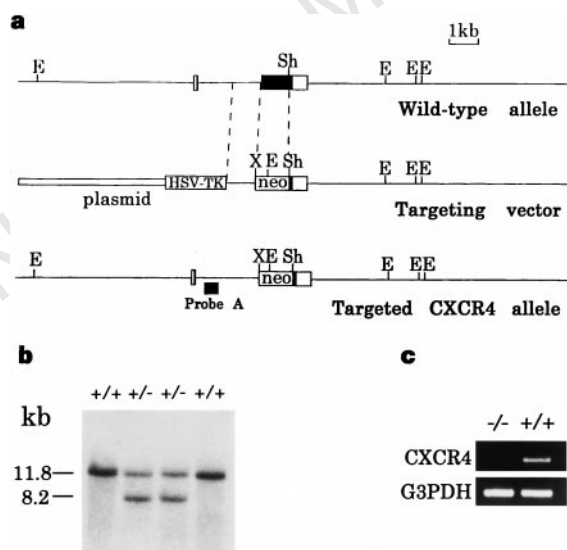


Figure 1 Targeted disruption of the CXCR4 gene. **a**, Targeting strategy. Wild-type CXCR4 locus (top), targeting vector (middle), and predicted mutant locus (bottom). The coding regions of the gene are indicated as filled boxes; empty boxes represent the 5' and 3' untranslated regions. Probe A is the external probe for Southern blot analysis. Restriction sites: E, *Eco*RI; Sh, *Sph*I; X, *Xho*I. **b**, Representative Southern blot analysis of tail DNA from wild-type (+/+) and heterozygous mutant (+/-) mice. The 11.8-kb wild-type and 8.2-kb targeted allele *Eco*RI-*Eco*RI fragments identified by probe A are indicated. **c**, RT-PCR analysis of CXCR4 expression. RT-PCR was also done for the G3PDH mRNA as a control for the presence of amplifiable RNA.

most of the branches in CXCR4^{-/-} embryos were single (Fig. 2d). SMA and SMV in the mesenteries of CXCR4^{-/-} embryos appeared to be normal (data not shown). Histological analysis by electron microscopy revealed that the ultrastructure of the endothelial and periendothelial cells of the aberrant mesenteric vessels in the E13.5 mutants was normal (Fig. 2e, f). In E17.5 wild-type embryos, whole-mount analysis showed that the large mesenteric vessels split into many branches and reach one side of the intestine (Fig. 2g, i). However, in CXCR4^{-/-} embryos these were almost absent and there were fewer large vessels with aberrant branching (Fig. 2h, j). Although this defective vasculature seemed to support organogenesis, we found that most E16.5 mutant embryos had multiple haemorrhages and/or congestion in the small intestine, probably as a result of circulatory disturbance of the organ (Fig. 2k). Thus, CXCR4 is responsible for the normal vascularization of the small intestine, leading to mesenteric vascular branching and/or remodelling.

In the stomach, large vessels arising from mesenchymal vessels along the lesser curvature were formed and distributed to the ventral and dorsal surfaces in wild-type embryos by E13.5 (Fig. 3a, c). The large vessels in E15.5 wild-type mice contained arteries and veins (Fig. 3c, inset). To our surprise, these large vessels were absent in all mutant embryos ($n = 12$; Fig. 3b, d). Formation of the small vascular network that surrounds the stomach seemed to be normal in the mutant embryos (Fig. 3c, d).

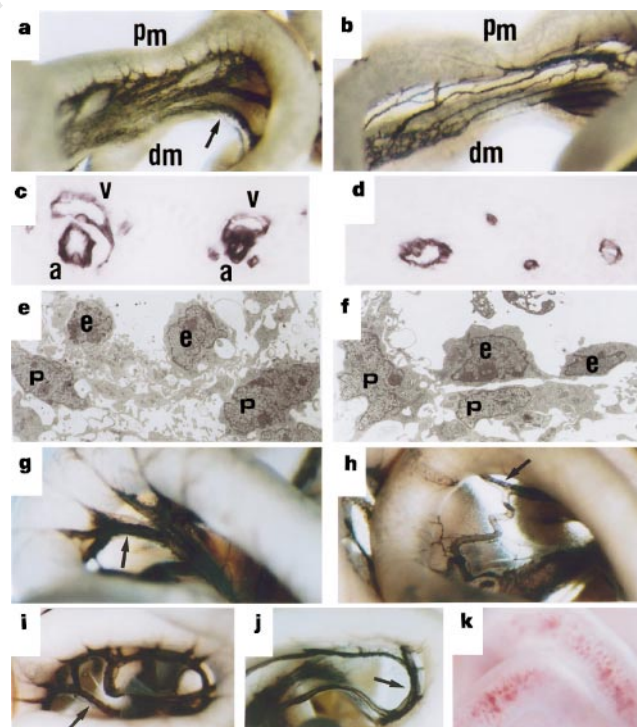


Figure 2 Formation of large vessels supplying intestines in embryonic mice. Whole-mount immunohistostaining of wild-type (**a, c, g, i**) and CXCR4^{-/-} (**b, d, h, j**) mesenteries and intestines with antibody against PECAM-1: mesentery and midgut loop region at E13.5 (**a, b**); second loop of jejunum at E17.5 (**g, h**); third loop of jejunum at E17.5 (**i, j**). (**c, d**), Cross-sections of stained mesenteries at E13.5. Branches of SMA or SMV are shown. (**e, f**), Ultrastructural analysis of wild-type and mutant vessels in mesentery at E13.5. **k**, Haemorrhagic unstained intestines at E16.5 mutants. Arrows in **a, g, i** indicate large branches of SMA or SMV supplying the small intestine in wild-type mesentery. Arrows in **h, j** indicate large vessels displaying aberrant running and/or branching in mutant mice. pm, Proximal part of midgut loop; dm, distal part of midgut loop; a, artery; v, vein; e, endothelial cell; p, periendothelial cell.

The stomachs and intestines of E18.5 mutant embryos had no obvious abnormalities in organogenesis (data not shown). These findings indicate that the vascular defects in CXCR4^{-/-} mice are not a secondary outcome of gastrointestinal tract organogenesis. Similar defects in the formation of particular vessels were observed in all mice lacking PBSF/SDF-1 (*n* = 9; Fig. 4a, b).

In E12.5 wild-type embryos, CXCR4 transcripts were found by *in situ* hybridization in endothelial cells of blood vessels in mesentery and in the wall of the intestine and stomach, with strong expression in branches arising from the SMA (Fig. 5b, e). The location and distribution of the CXCR4-positive cells was typical of endothelial cells but not of vascular wall (Fig. 5f, g). At E15.5, the amount of CXCR4 mRNA decreased and only faint signals were detected in the endothelial cells of vessels in mesenteries (data not shown). In contrast, at E12.5–15.5, PBSF/SDF-1 is expressed at high levels in mesenchymal cells surrounding the endothelial cells in the mesentery but not in endothelial cells or the walls of intestines and stomach (Fig. 3c, and data not shown). The expression patterns indicate that PBSF/SDF-1 produced by mesenchymal cells provides the paracrine signal that is received by CXCR4 on endothelial cells and plays a crucial role in the mesenteric mesenchyme. Thus the deficits in large vessels in the stomachs of CXCR4^{-/-} and PBSF/SDF-1^{-/-} embryos may result from defective vascular branching and/or remodelling in the mesenteric mesenchyme along the lesser curvature of the stomach.

We examined the vasculature of other organs by staining whole-mount yolk sac, brain and heart with monoclonal antibody against PECAM. We found no obvious difference between CXCR4^{-/-}, PBSF/SDF-1^{-/-} and wild-type embryos in the formation of large and small vessels in the yolk sac (E12.5, E14.5), head region (E11.5), or developing heart (E12.5–14.5; data not shown). *In situ* hybridization analysis revealed that CXCR4 was expressed in endothelial cells of the vessels in brain and heart (E12.5), although there was no

expression of PBSF/SDF-1 in the cells surrounding the endothelial cells (data not shown).

The number of B-cell progenitors in fetal livers of CXCR4^{-/-} mice was severely reduced (Fig. 6a) and developing haematopoietic cells found in wild-type bone marrow were absent in mutant bone marrow (Fig. 6b, c). In addition, defects of the membranous portion of the cardiac ventricular septum were found in E18.5 mutant embryos (Fig. 6e, f). These abnormalities are similar to those found in mice lacking PBSF/SDF-1, supporting the idea that CXCR4 is a primary physiological receptor for PBSF/SDF-1. In the bone marrow of E18.5 wild-type mice, CXCR4 transcripts were expressed in haematopoietic cells but not in the spindle-shaped stromal cells in which PBSF/SDF-1 transcripts are expressed (Fig. 6d). These expression patterns indicate that paracrine signalling may operate in bone marrow.

We have shown that CXCR4 and PBSF/SDF-1 are responsible for the formation of a mature vascular system to supply the gastrointestinal tract, probably by regulating vascular branching and/or remodelling processes in endothelial cells. Mutant analysis of receptor tyrosine kinases and their ligands (such as Flk-1, Flt-1, Tie-1, Tie-2, VEGF, angiopoietin-1 and PDGF-B) have indicated that these are critical for vascular development and that many are required for early and general blood-vessel development, including that of the extra-embryonic vasculature, such as the yolk sac^{1–3,12–18}. In contrast, the functions of CXCR4 and PBSF/SDF-1 are later, intra-embryonic and relatively organ-specific. Tie-2 and its ligand angiopoietin-1 seem to be responsible for branching and/or remodelling in the early vascular system^{15,18}. However, their roles in the formation of the mature vascular system in the gastrointestinal tract have not been determined, and angiopoietin-1^{-/-} mice lack microvascular pericytes, in contrast to CXCR4^{-/-} mice. The CXCR4 chemokine, PF4 (ref. 19), IL-8 (ref. 20), IP-10 (ref. 21) and Groβ (ref. 22) are also angiogenic regulators but their receptors in endothelial cells

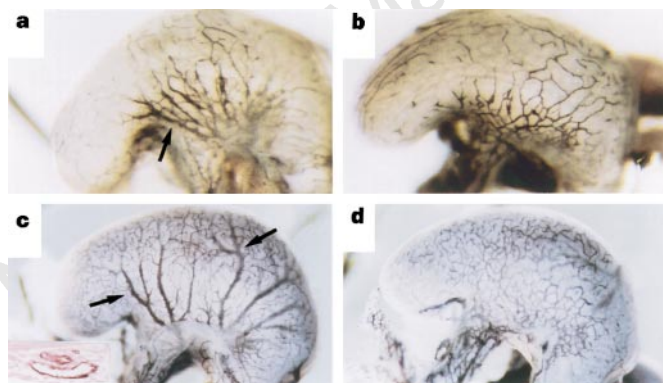


Figure 3 Vasculature of the stomach of embryonic mice. Whole-mount immunohistostaining of wild-type (a, c) and CXCR4^{-/-} (b, d) stomachs with antibody against PECAM-1 at E13.5 (a, b) and at E15.5 (c, d). Inset in c shows a haematoxylin-and-eosin-stained section of large vessels in the wall of stained stomach at E15.5. Arrows in a and c indicate large vessels not seen in mutant embryos (b, d).

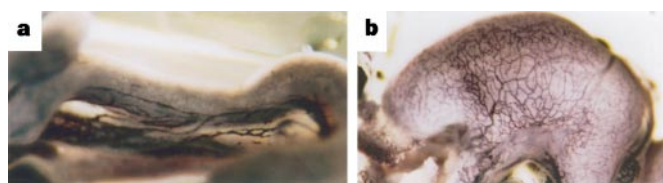


Figure 4 Vascular defects in the gastrointestinal tract of PBSF/SDF-1^{-/-} embryos. Whole-mount immunohistostaining, with antibody against PECAM-1, of PBSF/SDF-1^{-/-} mesentery and mid-gut loop region at E13.5 (a) and stomach at E15.5 (b).

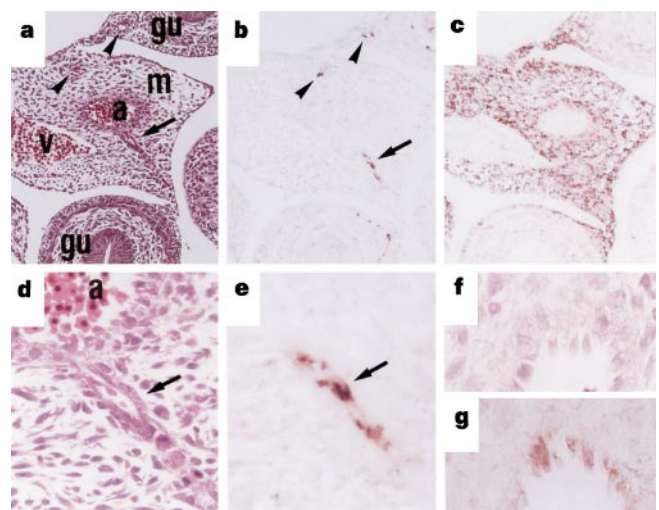


Figure 5 *In situ* hybridization analysis of CXCR4 and PBSF/SDF-1 expression in the gastrointestinal tract. Serial sections of wild-type E12.5 mesentery connecting to the mid-gut loop were stained with haematoxylin and eosin (a, d) and hybridized with probes specific for CXCR4 (b, e) and PBSF/SDF-1 (c, f). d, e, Enlargement of a vessel arising from SMA of a and b, showing expression of CXCR4 in endothelial cells of the vessel. Arrows and arrowheads in b and e, indicate the staining endothelial cells of blood vessels in the mesentery. Note the expression of PBSF/SDF-1 in mesenchymal cells surrounding endothelial cells in the mesentery (c). (f, g) Serial sections of the blood vessel with surrounding periendothelial cells in the E13.5 mesentery were stained with haematoxylin-and-eosin (f) and hybridized with CXCR4-specific probes (g). CXCR4 is expressed in endothelial cells but not in surrounding periendothelial cells. m, Mesentery; gu, midgut; a, SMA; v, SMV.

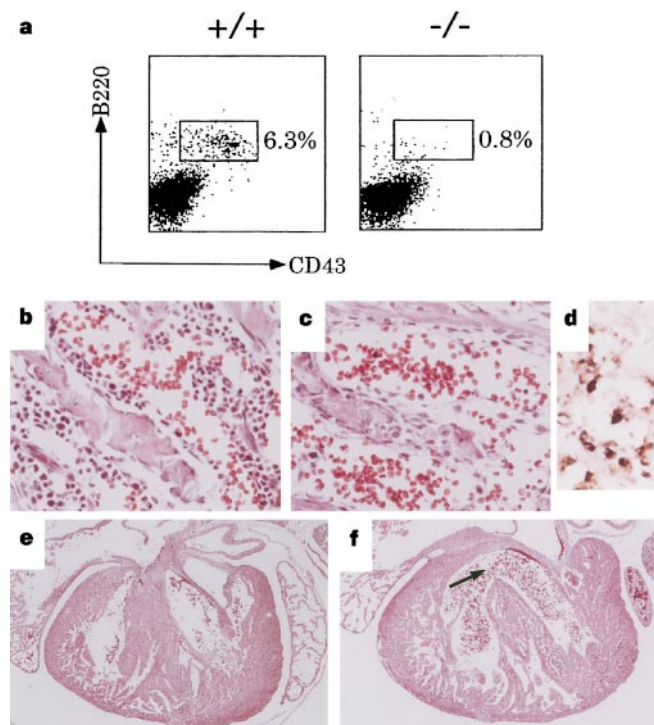


Figure 6 Defects of haematopoiesis and cardiogenesis in $CXCR4^{-/-}$ mice. **a**, Flow-cytometric analysis of E18.5 fetal liver cells stained with antibodies against B220 and antibodies against CD43. **b, c**, Sections of wild-type and mutant E18.5 bone marrows. **d**, Section of wild-type bone marrow, showing expression of CXCR4 in haematopoietic cells but not in spindle-shaped stromal cells. **e, f**, Sections of wild-type and mutant E18.5 hearts. Arrow in **f** indicates the ventricular septal defect.

and their physiological roles have not yet been defined. Although coagulation factor V (ref. 23) and tissue factor²⁴ are responsible for yolk-sac vasculature, the molecular signalling processes through which they exert their effects have not been identified. We have found a signalling system, including a chemokine and a G-protein-coupled receptor pair, that is crucial for vascular formation. Mice lacking an α -subunit of the heterotrimeric GTP-binding protein $G\alpha_{13}$ have a disorganized yolk-sac vascular system and enlarged and defective small vessels as embryos²⁵. Although the phenotypes are different from those in our $CXCR4$ -deficient mice, it should be tested whether $CXCR4$ couples to $G\alpha_{13}$.

$CXCR4$ and $CCR5$ are essential co-receptors for T-cell-line-tropic and macrophage-tropic HIV-1 infection, respectively^{8,26}. People homozygous for $CCR5$ deletion are resistant to HIV-1 infection and have no obvious health problems^{27–29}. However, our finding that mice lacking $CXCR4$ die *in utero* indicates that homologous defects in human $CXCR4$ are unlikely to be revealed, suggesting that additional genetic factors or homologous viable $CXCR4$ mutations may exist in long-term non-progressors resistant to T-cell-line-tropic HIV-1. □

Methods

Construction and analysis of $CXCR4^{-/-}$ mice. $CXCR4^{-/-}$ embryonic stem (ES) cells and mice were generated as described⁷. A 1.1-kb genomic fragment encompassing the 5' coding region of exon 2 of the murine $CXCR4$ gene was replaced by *neo*. For Southern blot analysis, DNA extracted from mouse tails was digested with *EcoRI* and hybridized to the 550-bp probe A which flanks the 5' homology region. RT-PCR was done using 3 μ g total RNA isolated from the fetal liver of E18.5 embryos. The 630-bp products were amplified using $CXCR4$ -specific primers. Primer sequences were as follows: forward, 5'-TAGCGG CCGCGTTGCCATGGAACCGAT-3'; reverse 5'-

GGCTCGACTTTGCATAA GGGTTAGCTG-3'. Histological analysis and flow cytometry analysis were done as described⁷.

Whole-mount immunohistostaining of embryos and organs. Whole-mount immunohistostaining was carried out as described³⁰.

Electron microscope. Organs were fixed in 2.5% glutaraldehyde/2% paraformaldehyde in cacodylate buffer, pH 7.4, then treated with 1% cacodylate-buffered (0.1 M) OsO_4 , dehydrated, embedded in Epon812 and sectioned on an Ultracut E ultramicrotome (Reichert-Jung). Ultrathin sections were stained with saturated uranyl acetate and lead citrate and examined in a Hitachi H7100 electron microscope.

PBSF/SDF-1 $^{-/-}$ mice. PBSF/SDF-1 mice were bred from heterozygote parents and genotyped by PCR as described⁷.

In situ hybridization. *In situ* hybridization was carried out as described⁷ using antisense transcription of murine CXCR4 or PBSF/SDF-1 cDNA fragments as probes.

Received 12 December 1997; accepted 6 April 1998.

1. Risau, W. Mechanisms of angiogenesis. *Nature* **386**, 671–674 (1997).
2. Folkman, J. & D'Amore, P. A. Blood vessel formation: what is its molecular basis? *Cell* **87**, 1153–1155 (1996).
3. Lindahl, P., Johansson, B. R., Lvevee, P. & Betsholtz, C. Pericyte loss and microaneurysm formation in PDGF-B-deficient mice. *Science* **277**, 242–245 (1997).
4. Bleul, C. C. *et al.* The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry. *Nature* **382**, 829–833 (1996).
5. Oberlin, E. *et al.* The CXC chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1. *Nature* **382**, 833–835 (1996).
6. Nagasawa, T. *et al.* Molecular cloning and characterization of a murine pre-B-cell growth-stimulating factor/stromal cell-derived factor 1 receptor, a murine homolog of the human immunodeficiency virus 1 entry coreceptor fusin. *Proc. Natl. Acad. Sci. USA* **93**, 14726–14729 (1996).
7. Nagasawa, T. *et al.* Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXC chemokine PBSF/SDF-1. *Nature* **382**, 635–638 (1996).
8. Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. HIV-1 entry coreceptor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**, 872–877 (1996).
9. Baggiolini, M., Dewald, B. & Moser, B. Human chemokines: an update. *Annu. Rev. Immunol.* **15**, 675–705 (1997).
10. Vecchi, A. *et al.* Monoclonal antibodies specific for endothelial cells of mouse blood vessels. Their application in the identification of adult and embryonic endothelium. *Eur. J. Cell Biol.* **63**, 247–255 (1994).
11. Baldwin, H. S. *et al.* Platelet endothelial cell adhesion molecule-1 (PECAM-1/CD31): alternatively spliced, functionally distinct isoforms expressed during mammalian cardiovascular development. *Development* **120**, 2539–2553 (1994).
12. Shalaby, F. *et al.* Failure of blood-island formation and vasculogenesis in Flk-1-deficient mice. *Nature* **376**, 62–66 (1995).
13. Fong, G.-H., Rossant, J., Gertsenstein, M. & Breitman, M. L. Role of the Flt-1 receptor tyrosine kinase in regulating the assembly of vascular endothelium. *Nature* **376**, 66–70 (1995).
14. Dumont, D. J. *et al.* Dominant-negative and targeted null mutations in the endothelial receptor tyrosine kinase, tek, reveal a critical role in vasculogenesis of the embryo. *Genes Dev.* **8**, 1897–1909 (1994).
15. Sato, T. N. *et al.* Distinct roles of the receptor tyrosine kinases Tie-1 and Tie-2 in blood vessel formation. *Nature* **376**, 70–74 (1995).
16. Carmeliet, P. *et al.* Abnormal blood vessel development and lethality in embryos lacking a single VEGF allele. *Nature* **380**, 435–439 (1996).
17. Ferrara, N. *et al.* Heterozygous embryonic lethality induced by targeted inactivation of the VEGF gene. *Nature* **380**, 439–442 (1996).
18. Suri, C. *et al.* Requisite role of angiopoietin-1, a ligand for the TIE2 receptor, during embryonic angiogenesis. *Cell* **87**, 1171–1180 (1996).
19. Maione, T. E. *et al.* Inhibition of angiogenesis by recombinant human platelet factor-4 and related peptides. *Science* **247**, 77–79 (1990).
20. Koch, A. E. *et al.* Interleukin-8 as a macrophage-derived mediator of angiogenesis. *Science* **258**, 1798–1801 (1992).
21. Luster, A. D., Greenberg, S. M. & Leder, P. The IP-10 chemokine binds to a specific cell surface heparan sulfate site shared with platelet factor 4 and inhibits endothelial cell proliferation. *J. Exp. Med.* **182**, 219–231 (1995).
22. Cao, Y., Chen, C., Weatherbee, J. A., Tsang, M. & Folkman, J. α - β , a CXC chemokine, is an angiogenesis inhibitor that suppresses the growth of Lewis lung carcinoma in mice. *J. Exp. Med.* **182**, 2069–2077 (1995).
23. Cui, J., Sue O'Shea, K., Purkayastha, A., Saunders, T. L. & Ginsburg, D. Fatal haemorrhage and incomplete block of embryogenesis in mice lacking coagulation factor V. *Nature* **384**, 66–68 (1996).
24. Carmeliet, P. *et al.* Role of tissue factor in embryonic blood vessel development. *Nature* **383**, 73–75 (1996).
25. Offermanns, S., Mancino, V., Revel, J.-P. & Simon, M. I. Vascular system defects and impaired cell chemokinesis as a result of $G\alpha_{13}$ deficiency. *Science* **275**, 533–535 (1997).
26. Fauci, A. S. Host factors and the pathogenesis of HIV-induced disease. *Nature* **384**, 529–535 (1996).
27. Liu, R. *et al.* Homozygous defect in HIV-1 coreceptor accounts for resistance of some multiply-exposed individuals to HIV-1 infection. *Cell* **86**, 367–377 (1996).
28. Samson, M. *et al.* Resistance to HIV-1 infection in caucasian individuals bearing mutant alleles of the R-5 chemokine receptor gene. *Nature* **382**, 722–725 (1996).
29. Dean, M. *et al.* Genetic restriction of HIV-1 infection and progression to AIDS by a detection allele of the $CCR5$ structural gene. *Science* **273**, 1856–1861 (1996).
30. Adachi, S., Yoshida, H., Kataoka, H. & Nishikawa, S.-I. Three distinctive steps in Peyer's patch formation of murine embryo. *Int. Immunol.* **9**, 507–514 (1997).

Acknowledgements. We thank K. Morihana, T. Ohito and T. Iwaki for technical assistance; T. Tanaka for technical advice; and T. Nakajima for discussion. This work was supported by grants from the Ministry of Education of Japan.

Correspondence and requests for materials should be addressed to T.N. (e-mail: immunol@osk.threewebnet.or.jp).

Function of the chemokine receptor CXCR4 in haematopoiesis and in cerebellar development

Yong-Rui Zou*, Andreas H. Kottmann†, Masahiko Kuroda*, Ichiro Taniuchi*‡ & Dan R. Littman*‡

* Division of Molecular Pathogenesis and ‡ Howard Hughes Medical Institute, Skirball Institute of Biomolecular Medicine, New York University Medical Center, New York, New York 10016, USA

† Department of Biochemistry and Molecular Biophysics, Columbia University College of Physicians and Surgeons, New York, New York 10032, USA

Chemokines and their receptors are important in cell migration during inflammation¹, in the establishment of functional lymphoid microenvironments², and in organogenesis³. The chemokine receptor CXCR4 is broadly expressed in cells of both the immune and the central nervous systems^{4,5} and can mediate migration of resting leukocytes and haematopoietic progenitors in response to its ligand, SDF-1 (refs 6–9). CXCR4 is also a major receptor for strains of human immunodeficiency virus-1 (HIV-1) that arise during progression to immunodeficiency and AIDS dementia¹⁰. Here we show that mice lacking CXCR4 exhibit haematopoietic and cardiac defects identical to those of SDF-1-deficient mice³, indicating that CXCR4 may be the only receptor for SDF-1. Furthermore, fetal cerebellar development in mutant animals is markedly different from that in wild-type animals, with many proliferating granule cells invading the cerebellar anlage. This is, to our knowledge, the first demonstration of the involvement of a G-protein-coupled chemokine receptor in neuronal cell migration and patterning in the central nervous system. These results may be important for designing strategies to block HIV entry into cells and for understanding mechanisms of pathogenesis in AIDS dementia.

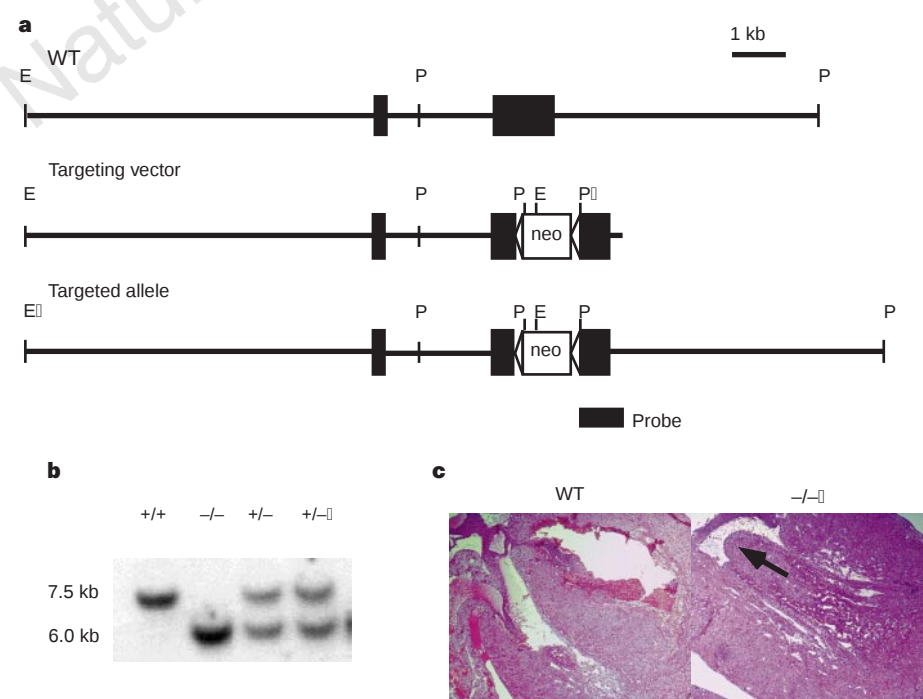
Disruption of the murine CXCR4 gene (Fig. 1a, b) resulted in fetal lethality in homozygous mutant animals. Rare homozygous mutant offspring were born alive and died within a few hours, although at

embryonic day (E) 13.5 CXCR4-deficient embryos were macroscopically indistinguishable from their littermates (data not shown). By E17.5, however, half of the CXCR4^{-/-} embryos had died. The viable E17.5 CXCR4^{-/-} embryos normally had reduced body size, with a mass averaging only 74% of their wild-type (+/+ or +/-) littermates. These embryos were still able to open their mouths and had flexion of the neck and body, indicating that the muscular-neuronal system controlling global stereotyped movements had developed well. At E17.5, more than 70% of the moribund CXCR4^{-/-} embryos had developed generalized oedema, and, on histological analysis, all mutant embryos exhibited dysplasia of the ventricular septum (Fig. 1c), a defect identical to the cardiac defect seen previously in SDF-1^{-/-} mice³.

SDF-1^{-/-} mice have defective B-cell lymphopoiesis and severely impaired bone-marrow myelopoiesis, despite normal myeloid development in the fetal liver³. We observed similar defects in the CXCR4^{-/-} mice, as we could not detect cells expressing B220 or CD43 in fetal liver or bone marrow (not shown). In the fetal liver of the CXCR4^{-/-} mice, cells of the myeloid lineage, including macrophages (CD11b⁺), granulocytes, monocytes (Gr1⁺, CD11b⁺) and megakaryocytes (CD61⁺), developed normally like their wild-type littermates. However, the cellularity of these cell types was markedly reduced in the bone marrow of the mutant animals (Fig. 2a). These results, together with the identical defect in cardiac development in both CXCR4^{-/-} and SDF-1^{-/-} embryos, indicate that CXCR4 may be the only physiological receptor for SDF-1 during fetal development. This conclusion is further supported by the finding that fetal liver cells from mutant mice failed to migrate towards SDF-1 in an *in vitro* transwell chemotaxis assay (Fig. 2b).

The absence of B220⁺ CD43⁺ cells in the mutant mice indicates that B-cell lymphopoiesis is blocked before the pro-B-cell stage¹¹. Possible causes of the absence of pro-B cells in both CXCR4^{-/-} and SDF-1^{-/-} mice include failure in fate determination and proliferation of B-lineage precursors, or impaired homing of B-cell progenitors to a supportive niche. We used an *in vitro* clonal assay^{12,13} to distinguish between these two possibilities. As shown in Fig. 2c, no substantial pro-B-cell clones were generated from cells of mutant mice on cultivation of mutant fetal liver cells with the S17 stromal cell line. As progenitors are placed directly in a supportive environment in this assay, the problem of cell homing is circumvented. Therefore, the absence of B-lineage cells in CXCR4^{-/-} mice seems to

Figure 1 Targeted disruption of the CXCR4 gene. **a**, Restriction maps of the wild-type (WT) allele, targeting vector, and targeted allele are shown. E, EcoRI; P, PstI; neo, neomycin-resistant gene. **b**, Southern blot analysis of placental DNA of E17.5 embryos obtained from heterozygous intercrosses. DNA was digested with PstI and hybridized with the diagnostic probe as indicated. **c**, Haematoxylin-and-eosin-stained sagittal sections of hearts in E17.5 wild-type and mutant (-/-) embryos. The arrow points to the ventricular septal defect.



result from deficiency in commitment and/or proliferation of B-progenitor cells.

CXCR4 is highly expressed in the thymus, particularly in immature $CD4^+ CD8^+$ cells⁴. However, T lymphopoiesis occurred normally in the mutant embryos at E17.5 (not shown). At this stage, most thymocytes are immature $CD4^+ CD8^+$ cells; few $CD4^+ CD8^-$ or $CD4^- CD8^+$ mature T lymphocytes are present, and there is little or no emigration of thymocytes to the peripheral lymphoid organs. Emigration commences only after birth through a process in which the G_i proteins are important^{14,15}. As CXCR4 expressed in resting T cells is coupled to G_{i2} (refs 4, 10) and as mature T cells can migrate in response to SDF-1 stimulation⁹, migration and recirculation of mature T cells might require signals delivered through CXCR4. To determine whether maturation and emigration of thymocytes

requires CXCR4, we engrafted the thymuses from E17.5 CXCR4^{-/-} embryos under the kidney capsule of mice lacking T-cell antigen receptor (TCR)- α . As the recipient mice are deficient in $\alpha\beta$ -T-cell development, any TCR- $\alpha\beta$ T cells detected in the periphery are derived from the donor thymus. Our results indicate that mature $CD4^+ CD8^-$ and $CD4^- CD8^+$ T cells developed normally in the implanted thymuses. Moreover, these T cells efficiently populated the peripheral lymphoid organs, including peripheral blood (Fig. 2d), spleen, and mesenteric lymph nodes (results not shown). Thus, CXCR4 has a negligible effect on thymocyte maturation and subsequent migration to lymphoid organs.

As CXCR4 messenger RNA is expressed in brain⁴, we used *in situ* hybridization to study the embryonic mouse brain at E13, E15 and E18. CXCR4 mRNA transcripts were found in many regions of the

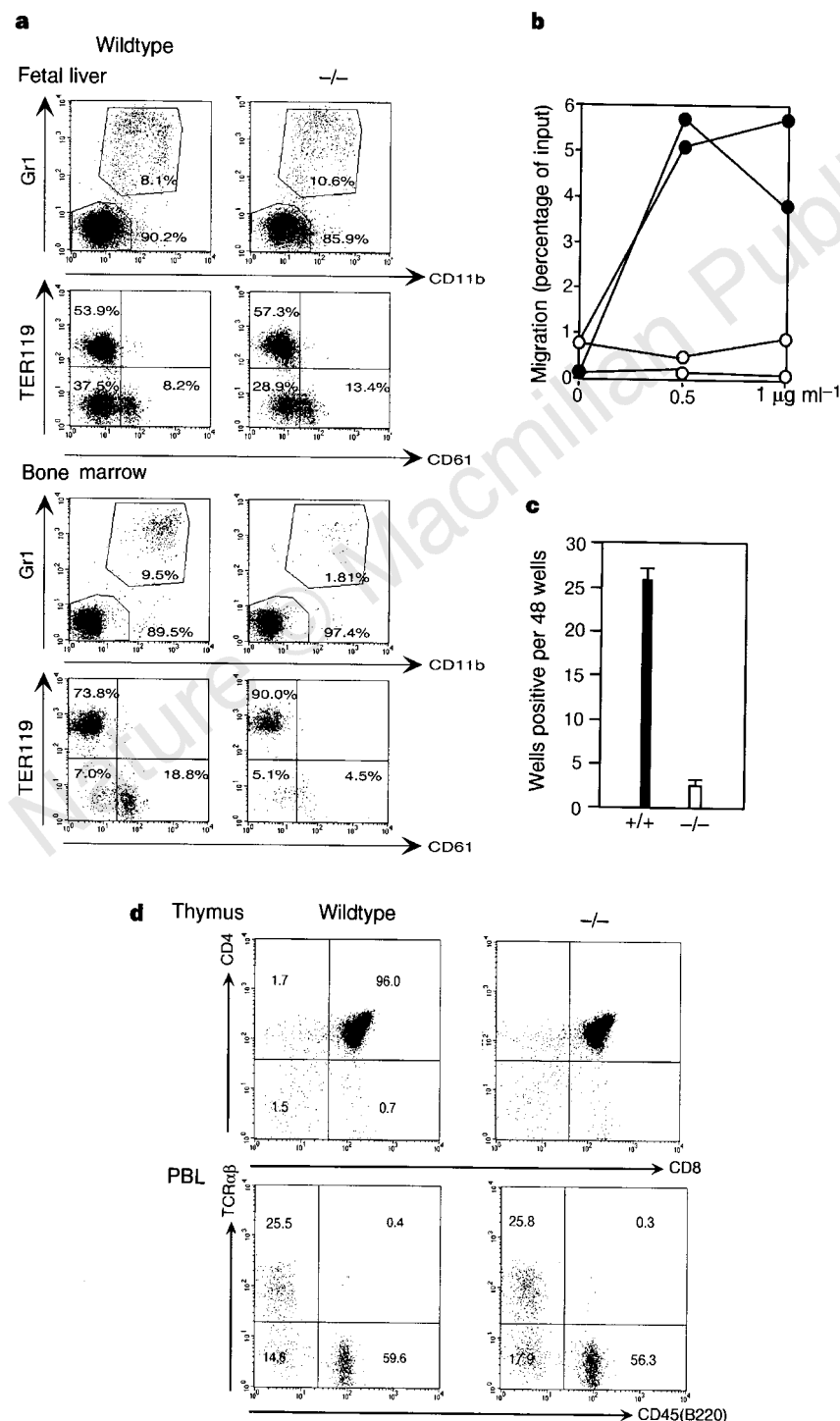


Figure 2 Impaired haematopoiesis and SDF-1-induced chemotaxis, but normal development of T cells, in CXCR4^{-/-} mice. **a**, Flow-cytometric analysis of fetal liver cells and bone-marrow cells isolated from E17.5 embryos. Cells were stained with antibodies against the indicated cell-surface markers (Gr1, CD11b, CD61 and TER119), and the percentages of cells in the marked gates are indicated. **b**, Transwell assay measurement of SDF-1-induced chemotaxis of fetal liver cells from E17.5 mutant embryos (open circles) and wild-type littermates (filled circles). Flow-cytometric analysis showed that >90% of cells migrating into the bottom chamber expressed myeloid-lineage markers. Results from two independent experiments are shown. **c**, *In vitro* clonal assay of pro-B-cell colony formation. Black bar, wild-type; white bar, CXCR4^{-/-}. Results represent the mean of four separate experiments. **d**, Flow-cytometric analysis of thymocytes isolated from transplanted thymi of mutant and control embryos. B and T lymphocytes from the peripheral blood of the recipient mice were detected using antibodies against the lineage-specific markers CD45 (B220) and TCR- $\alpha\beta$. The percentage of cells in the given quadrants is indicated.

developing brain, including the retina, olfactory epithelium, olfactory bulb, hippocampus, cerebellum and spinal cord (Fig. 3, and data not shown). A similar pattern of CXCR4 mRNA expression was found in rat brain⁵. Expression was observed in regions of the cerebellum enriched in proliferating cells, including neuroepithelium, rhombic lip and the external granule layer (EGL) (Fig. 3). We then compared brains of the mutant mice with those of wild-type littermates. Grossly, the structures of brains from CXCR4^{-/-} mice were comparable with those from wild-type animals (data not shown). However, closer examination showed that the laminar structure of the cerebellum of CXCR4^{-/-} mice was aberrant. Although the EGL was present, clusters of granule cells were found in ectopic positions, beneath the Purkinje cell layer or intermingled with Purkinje cells (Fig. 4a, b). These abnormally placed cells expressed the Math1 protein, a marker of external granule cells¹⁶ (Fig. 4c, d). All 20 mutant cerebella examined showed normal patterning of neuronal layers in the caudal part close to the rhombic lip, indicating that the defect in granule cell migration is probably a result of premature migration from the EGL rather than of abnormal laminar migration from the rhombic lip. The descending migration of EGL cells to form the internal granule layer (IGL) is normally a postmitotic event that occurs primarily after birth. In the mutant mice, migration of the granule cells from the EGL was observed as early as E17.5, and involved cells that continued to proliferate, as assessed by incorporation of bromodeoxyuridine during a 2-h pulse labelling (Fig. 4e, f). In contrast to the granule cells, the Purkinje cells were situated correctly in mutant embryos, as shown by staining with an anti-calbindin antibody (not shown).

The perpendicular migration of EGL cells may be directed by a

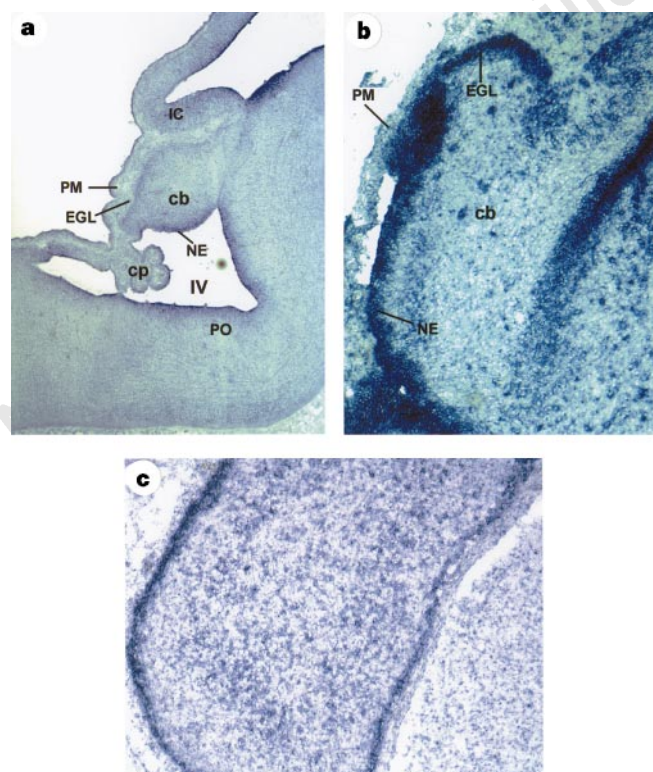


Figure 3 Localization of CXCR4 mRNA in the wild-type developing brain as shown by *in situ* hybridization. A sense-chain RNA probe used as a negative control gave no significant signal above background. **a**, Sagittal section showing CXCR4 expression (darker shading) in neuronal precursors on the surface of the cerebellar anlage of E13.5 embryo (original magnification, $\times 50$). **b**, Sagittal section showing increased CXCR4 expression in E15.5 cerebellum ($\times 100$). **c**, Transverse section of the lateral aspect of the E18.5 cerebellum (original magnification, $\times 100$). cb, Cerebellum; cp, choroid plexus; EGL, external granule layer; IV, fourth ventricle; IC, inferior colliculus; NE, neuroepithelium; PM, pia mater; PO, pons.

reciprocal interaction between the granule neurons and Bergmann cells (the cerebellar radial glial cells). Although the postmitotic granule cells undergo differentiation and induce Bergmann cell fibre extension, Bergmann cells provide a scaffold for neuronal migration and positioning^{17,18}. Therefore, the dislocation of EGL in the mutant embryos early in cerebellar development could be a consequence of the malformation of Bergmann cells. Staining with brain-specific lipid-binding protein (BLBP)¹⁹, which labels radial glial cells, showed that these cells were properly localized with their fibres arranged in a normal configuration and orientation in the CXCR4^{-/-} mice (Fig. 4g, h). In some mutant embryos, there was an increase in the number of neuronal cells along the Bergmann glial fibres (Fig. 4h).

These results indicate that CXCR4-mediated signalling is required to prevent premature migration of proliferating granule cells inwards from the EGL. This could involve the induction of

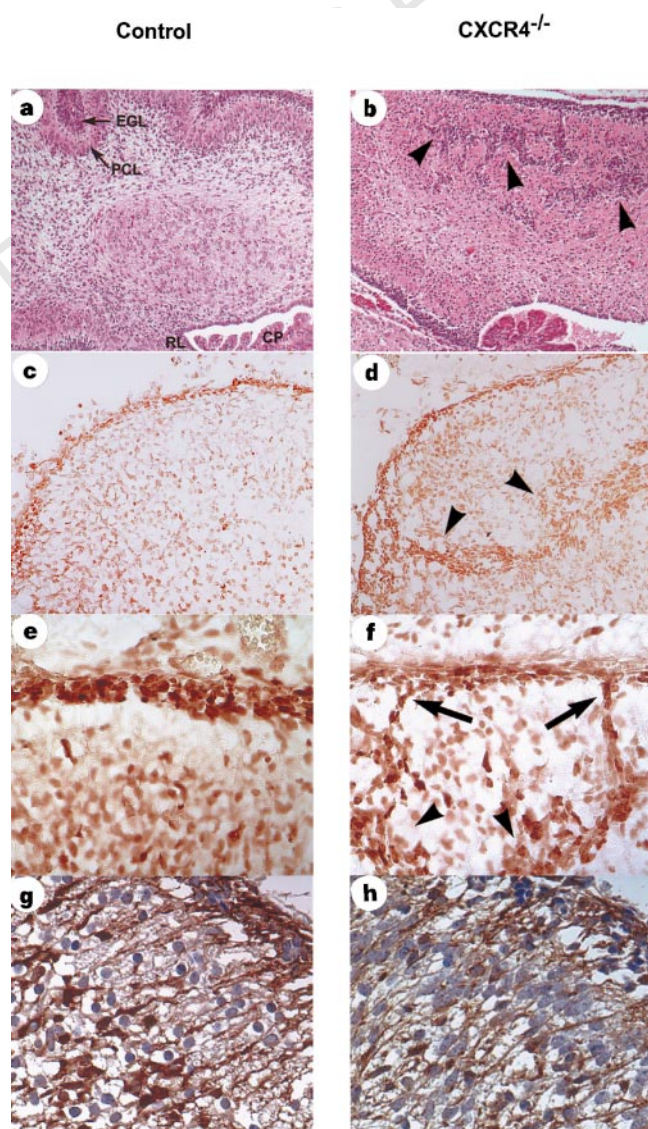


Figure 4 Abnormal migration of cerebellar EGL cells in CXCR4^{-/-} embryos. **a, b**, Haematoxylin-and-eosin-stained sagittal sections, showing dislocated cell aggregates underneath the Purkinje cell layer (arrow heads) in postnatal day 0 mutant animals. **c-h**, Sagittal sections prepared from E17.5 embryos. Sections were stained with antibodies against Math1 (**c, d**), BrdU (**e, f**) or BLBP (**g, h**; counterstained with haematoxylin). Arrowheads indicate ectopic granule cells. Migrating granule cells from the EGL are viewed by higher magnification (arrows, **f**). CP, choroid plexus; PCL, Purkinje cell layer; RL, rhombic lip. Anterior is to the right and dorsal to the top in each photograph. Original magnifications: **a, b**, $\times 100$; **c, d**, $\times 200$; **e-h**, $\times 400$.

adhesive interactions that prevent inward cellular migration, similar to induction of increased integrin–ligand avidity in leukocyte–endothelium interactions²⁰. Alternatively, CXCR4-mediated signals may desensitize cells to prevent inward migration in response to different chemoattractants. Determining the mechanism involved will require further studies with isolated granule cells. It is unknown whether the ligand for CXCR4 in cerebellar development is SDF-1, as no similar cerebellar abnormality was reported in SDF-1-deficient mice³. As signalling through chemoattractant G-protein-coupled receptors results in reorganization of the cellular cytoskeleton and in polarized cell movement²¹, such molecules are good candidates for receptors involved in neuronal cell migration and in axon guidance in the developing nervous system. Further studies will be required to determine whether chemokines function like netrins and ephrins in providing guidance cues to axons^{22,23}.

Thus the chemokine receptor CXCR4 has several important functions in addition to inducing leukocyte chemotaxis. It is essential at the earliest stages of B-cell lymphopoiesis, in colonization of bone marrow by multipotential haematopoietic cells, in cardiac septum formation, and in cerebellar neuronal layer formation. The observation that CXCR4 acts in the development of both the immune system and the central nervous system may be important, as CXCR4 is a receptor for strains of HIV-1 that become prevalent with the onset of immunodeficiency and AIDS dementia^{24,25}. HIV infection may perturb CXCR4 function in the adult central nervous system, resulting in some of the neurological manifestations. Further investigation of the role of CXCR4 in the adult brain will be required to determine whether it acts in virus-induced neurodegenerative disease. Our results also show that SDF-1 and CXCR4 form a monogamous ligand–receptor pair during early haematopoietic and cardiac development. Blocking of CXCR4 with SDF-1 or with small molecule inhibitors interferes effectively with HIV entry^{6,7,26}. If CXCR4 also has non-redundant functions in adults, antagonists that block HIV entry may be harmful if they also interfere with chemokine binding. An understanding of the functions of this chemokine receptor will be essential for guiding the design of therapies aimed at blocking HIV entry into cells. □

Methods

Targeted disruption of CXCR4. The mouse CXCR4 gene was cloned from a 129/sv genomic DNA library (Stratagene). The targeting vector contained 8.6 kilobases (kb) of 5' and 0.8 kb of 3' homologous regions. The neomycin cassette flanked by LoxP sites²⁷, used as a positive selection marker, was inserted into the second exon of the CXCR4 gene at the *KpnI* site. The 3' homologous region was obtained by polymerase chain reaction (PCR) using primers as follows: 5' primer: (5'-GCCGTGACGCTACCTCGCCATTGTCCACG-3'), and 3' primer 5'-GGCATCGATGTACCTCTAGACAGTCTCTTATATCTGGAAAA TG-3'). Chimaeric mice from three independent embryonic stem cells lines transmitted the mutated CXCR4 gene into the germ line.

Histology, bromodeoxyuridine labelling, immunohistochemistry and *in situ* hybridization. Pregnant females were killed at indicated time points. Embryos were removed and kept in cold PBS, and placentas from corresponding embryos were collected for genotyping. Heads and hearts of the embryos were dissected, immersion-fixed in Bouin's solution at room temperature or 4% paraformaldehyde at 4°C, and then processed into paraffin section by routine procedure. Sagittal sections of 5 µm were stained with haematoxylin–eosin. To detect proliferating cells in the cerebellum, pregnant females were injected with bromodeoxyuridine (BrdU) (150 mg per kg body weight, intraperitoneally) and killed 2 h later. Embryos were isolated and fixed with 4% paraformaldehyde at 4°C, and 20-µm frozen sections were prepared. For immunohistochemical staining, sections were incubated with different antibodies overnight at 4°C. Primary antibodies used were as follows: anti-calbindin (1:3,000, SWant), anti-BLBP (1:5,000, a gift from N. Heitz), anti-Math1 (1:500, a gift from J. Johnson), anti-BrdU (1:1, Amersham). Antibody binding was detected using peroxidase-coupled anti-rabbit antibody (Boehringer, 1:100) or anti-mouse immunoglobulin (Sigma, 1:100). *In situ* hybridization using the digoxigenin system (Boehringer) was done as

described²⁸. The CXCR4 probe was generated by antisense transcription of the 5' 580-base-pair *Bam*HI complementary DNA fragment.

***In vitro* clonogenic assay and flow cytometry.** Single-cell suspension of fetal liver, bone marrow and thymus was prepared. A standard protocol was used for the detection of clonal pro-B cells^{12,13}. Briefly, 96-well plates were prepared by seeding 2,000 S17 cells²⁹ per well; after overnight culture, plates were irradiated (3,000 rads). Five hundred fetal liver cells were plated into each well at a final volume of 200 µl in the Opti-MEM medium (Gibco) supplemented with 15% FCS (Hyclone), 5 × 10⁻⁵ M 2-mercaptoethanol, and 20 U ml⁻¹ recombinant interleukin-7 (Gibco). The medium was changed every 4 days and lymphocyte clones were scored after 10–12 days. Cell identification was confirmed by fluorescence-activated cell sorting (FACS) analysis by staining with antibodies against CD45 (B220) (phycoerythrin (PE)-conjugated) and CD43 (fluorescein isothiocyanate (FITC)). Antibodies used to detect myeloid-lineage cells were anti-CD11b (FITC) and anti-Gr1 (PE); for the erythroid lineage, anti-TER119 (PE); for megakaryocytes, anti-CD61 (FITC); for T-lymphoid cells, anti-CD4 (FITC), anti-CD8 (PE) and anti-TCRαβ (FITC). All monoclonal antibodies used in the flow cytometry were from PharMingen.

Fetal thymus transplantation. Fetal thymuses were removed from embryos at E17.5. Recipient mice (TCR-α^{-/-}) were anaesthetized with Avertin solution. Two thymic lobes from each embryo were placed under the kidney capsule. Three or four weeks after transplantation, recipients were killed, and transplanted thymuses and lymphoid organs were dissected and analysed by flow cytometry.

***In vitro* cell migration assay.** 2 × 10⁵ fetal liver cells in 100 µl were loaded into each Transwell filter (3-µm pore filter Transwell, 24-well cell clusters, Costar). Filters were then plated in each well containing 600 µl medium supplemented with different concentrations of SDF-1 as indicated. After 3–4 h incubation at 37°C, the upper chambers were removed, the cells in the bottom chamber were collected and counted, and their identities were confirmed by flow cytometry.

Received 2 February; accepted 21 April 1998.

- Murphy, P. M. Chemokine receptors: structure, function and role in microbial pathogenesis. *Cytokine Growth Factor Rev.* **7**, 47–64 (1996).
- Förster, R. et al. A putative chemokine receptor, BLR1, directs B cell migration to defined lymphoid organs and specific anatomic compartments of the spleen. *Cell* **87**, 1037–1047 (1996).
- Nagasawa, T. et al. Defects of B-cell lymphopoiesis and bone-marrow myelopoiesis in mice lacking the CXCR4 chemokine PBSF/SDF-1. *Nature* **382**, 635–638 (1996).
- Moepf, B., Frod, R., Rodewald, H. R., Baggiolini, M. & Gierschik, P. Two murine homologues of the human chemokine receptor CXCR4 mediating stromal cell-derived factor 1α activation of G12 are differentially expressed *in vivo*. *Eur. J. Immunol.* **27**, 2102–2112 (1997).
- Jazin, E. E., Soderstrom, S., Ebendal, T. & Larhammar, D. Embryonic expression of the mRNA for the rat homologue of the fusin/CXCR-4 HIV-1 co-receptor. *J. Neuroimmunol.* **79**, 148–154 (1997).
- Oberlin, E. et al. The CXCR4 chemokine SDF-1 is the ligand for LESTR/fusin and prevents infection by T-cell-line-adapted HIV-1. *Nature* **382**, 833–835 (1996).
- Bluel, C. C. et al. The lymphocyte chemoattractant SDF-1 is a ligand for LESTR/fusin and blocks HIV-1 entry. *Nature* **382**, 829–833 (1996).
- Aiuti, A., Webb, I. J., Bluel, C., Springer, T. & Gutierrez-Ramos, J. C. The chemokine SDF-1 is a chemoattractant for human CD34⁺ hematopoietic progenitor cells and provides a new mechanism to explain the mobilization of CD34⁺ progenitors to peripheral blood. *J. Exp. Med.* **185**, 111–120 (1997).
- Bluel, C. C., Fuhlbrigge, R. C., Casasnovas, J. M., Aiuti, A. & Springer, T. A. A highly efficacious lymphocyte chemoattractant, stromal cell-derived factor 1 (SDF-1). *J. Exp. Med.* **184**, 1101–1109 (1996).
- Feng, Y., Broder, C. C., Kennedy, P. E. & Berger, E. A. HIV-1 entry cofactor: functional cDNA cloning of a seven-transmembrane, G-protein-coupled receptor. *Science* **272**, 872–877 (1996).
- Hardy, R. R., Carmack, C. E., Shinton, S. A., Kemp, J. D. & Hayakawa, K. Resolution and characterization of pro-B and pre-pro-B cell stages in normal mouse bone marrow. *J. Exp. Med.* **173**, 1213–1225 (1991).
- Hayashi, S. et al. Stepwise progression of B lineage differentiation supported by interleukin 7 and other stromal cell molecules. *J. Exp. Med.* **171**, 1683–1695 (1990).
- Delassus, S. & Cumano, A. Circulation of hematopoietic progenitors in the mouse embryo. *Immunity* **4**, 97–106 (1996).
- Chaffin, K. E. et al. Dissection of thymocyte signaling pathways by *in vivo* expression of pertussis toxin ADP-ribosyltransferase. *EMBO J.* **9**, 3821–3829 (1990).
- Rudolph, U. et al. Ulcerative colitis and adenocarcinoma of the colon in Gαi2-deficient mice. *Nature Genet.* **10**, 143–150 (1995).
- Akazawa, C., Ishibashi, M., Shimizu, C., Nakanishi, S. & Kageyama, R. A mammalian helix-loop-helix factor structurally related to the product of *Drosophila* proneural gene *atonal* is a positive transcriptional regulator expressed in the developing nervous system. *J. Biol. Chem.* **270**, 8730–8738 (1995).
- Hatten, M. E. The role of migration in central nervous system neuronal development. *Curr. Opin. Neurobiol.* **3**, 38–44 (1993).
- Hatten, M. E. & Heintz, N. Mechanisms of neural patterning and specification in the developing cerebellum. *Annu. Rev. Neurosci.* **18**, 385–408 (1995).
- Feng, L., Hatten, M. E. & Heintz, N. Brain lipid-binding protein (BLBP): a novel signaling system in the developing mammalian CNS. *Neuron* **12**, 895–908 (1994).
- Springer, T. A. Adhesion receptors of the immune system. *Nature* **346**, 425–434 (1990).
- Campbell, J. J. et al. Chemokines and the arrest of lymphocytes rolling under flow conditions. *Science* **279**, 381–384 (1998).
- Serafini, T. et al. Netrin-1 is required for commissural axon guidance in the developing vertebrate nervous system. *Cell* **87**, 1001–1014 (1996).
- Drescher, U., Bonhoeffer, F. & Müller, B. K. The Eph family in retinal axon guidance. *Curr. Opin. Neurobiol.* **7**, 75–80 (1997).
- Scarlatti, G. et al. *In vivo* evolution of HIV-1 co-receptor usage and sensitivity to chemokine-mediated suppression. *Nature Med.* **3**, 1259–1265 (1997).

25. Connor, R. I., Sheridan, K. E., Ceradini, D., Choe, S. & Landau, N. R. Change in coreceptor use correlates with disease progression in HIV-1-infected individuals. *J. Exp. Med.* **185**, 621–628 (1997).
26. Donzella, G. A. et al. AMD3100, a small molecule inhibitor of HIV-1 entry via the CXCR4 co-receptor. *Nature Med.* **4**, 72–77 (1998).
27. Gu, H., Zou, Y.-R. & Rajewsky, K. Independent control of immunoglobulin switch recombination at individual switch regions evidenced through Cre-loxP-mediated gene targeting. *Cell* **73**, 1155–1164 (1993).
28. Schaefer-Wiemers, N. & Gerfin-Moser, A. A single protocol to detect transcripts of various types and expression levels in neural tissue and cultured cells: *in situ* hybridization using digoxigenin-labelled cRNA probes. *Histochemistry* **100**, 431–440 (1993).
29. Landreth, K. S. & Dorshkind, K. Pre-B cell generation potentiated by soluble factors from a bone marrow stromal cell line. *J. Immunol.* **140**, 845–852 (1988).

Acknowledgements. We thank F. Hatan and M.-J. Sunshine for technical assistance; S. Vukmanovic for help with the thymic transplant experiments; J. Johnson and N. Heitz for anti-Math1 and anti-BLBP antibodies; K. Dorshkind and R. R. Hardy for the S17 cell line; and G. Fishell, A. Joyner, M. Chao, C. Mason, S. Jung, V. KewalRamani and C. Davis for comments on the manuscript; Y.-R.Z. thanks H. Gu for his continuous support. This work was supported by an NIH grant (to D.R.L.). Y.-R.Z. is the recipient of a postdoctoral fellowship from the Irvington Institute, D.R.L. is an Investigator of the Howard Hughes Medical Institute.

Correspondence and requests for materials should be addressed to Y.-R.Z. (e-mail: zou@saturn.med.nyu.edu).

Histone macroH2A1 is concentrated in the inactive X chromosome of female mammals

Carl Costanzi & John R. Pehrson

Department of Animal Biology, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

In female mammals one of the X chromosomes is rendered almost completely transcriptionally inactive^{1,2} to equalize expression of X-linked genes in males and females. The inactive X chromosome is distinguished from its active counterpart by its condensed appearance in interphase nuclei³, late replication⁴, altered DNA methylation², hypoacetylation of histone H4 (ref. 5), and by transcription of a large *cis*-acting nuclear RNA called *Xist*^{6–10}. Although it is believed that the inactivation process involves the association of specific protein(s) with the chromatin of the inactive X, no such proteins have been identified¹¹. We discovered a new gene family encoding a core histone which we called macroH2A (mH2A)^{12,13}. The amino-terminal third of mH2A proteins is similar to a full-length histone H2A, but the remaining two-thirds is unrelated to any known histones. Here we show that an mH2A1 subtype is preferentially concentrated in the inactive X chromosome of female mammals. Our results link X inactivation with a major alteration of the nucleosome, the primary structural unit of chromatin.

We examined the distribution of mH2A in mouse liver nuclei by immunofluorescence using antibodies against the non-histone region of one of the mH2A1 subtypes, mH2A1.2 (Fig. 1a). Most hepatocyte nuclei were brightly stained by these antibodies (Fig. 2a), although the nuclei of bile duct cells, endothelial cells and connective tissue showed less staining (data not shown). Speckled staining was present through most of the nuclei of both males and females. The nuclei of females, however, also had large, distinct mH2A-dense regions, which we name macrochromatin bodies (MCBs) (Fig. 2a, b). We found this sex difference in all mice we examined, including several sets of siblings, as well as in dog liver sections and human primary skin fibroblasts (data not shown). We detected MCBs using several methods of tissue fixation and a variety of polyclonal antibodies, including ones raised in rabbits or chickens and against the non-histone region of mH2A1.1 (data not shown). MCBs were usually perinucleolar, but little or no mH2A1.2 staining was detected in nucleoli, as assessed by double-label immunofluorescence with a monoclonal antibody against the nucleolar protein fibrillarin¹⁴ (data not shown).

To quantify the sex specificity of MCBs in liver parenchyma, we

identified complete nuclei in liver sections by confocal microscopy and assessed their MCB content (Fig. 2c). In mice, 85% of nuclei from females contained MCBs, compared with less than 1% of nuclei from males. A similar distribution was observed in dogs: 90% of female nuclei contained MCBs compared with less than 1% of male nuclei. The relative mH2A1 protein content of female and male mouse livers was similar (Fig. 1b), and previously estimated in rat liver to be one mH2A per 30 nucleosomes¹².

The female specificity of MCBs suggested a relationship to X-chromosome inactivation. We therefore localized MCBs relative to X chromosomes by staining MCBs in female mouse liver sections by immunofluorescence and then localizing the X chromosomes in these sections by fluorescent *in situ* hybridization (FISH) with a DNA probe that 'paints' mouse X chromosomes (X-paint)¹⁵ (Fig. 3a). Using confocal microscopy, we found that 99% of MCBs colocalized to an X chromosome and 43% of X chromosomes colocalized to an MCB (Fig. 3b). In a control experiment, MCBs never colocalized with chromosome 4 (data not shown). The X-paint results also confirmed that hepatocyte nuclei with more than one MCB (Fig. 2a, c) were polyploid (Fig. 3a). The incidence of polyploid nuclei was higher in older mice (data not shown).

These results indicate that the female-specific MCBs involve one of the two X chromosomes. To determine which X chromosome is involved, we examined the nuclear distribution of mH2A1 proteins in female mice with one X-chromosome, male mice with two and human Klinefelter fibroblasts with four. X/0 mice are females with just one X chromosome (the active X), as is found in human Turner's syndrome. The mH2A1.2 staining pattern of X/0 liver sections was identical to that of normal males (data not shown). Sex-reversed mice have one normal X and one (designated X_{scr}) that carries a translocated piece of the Y chromosome including the sex-determining locus¹⁶. XX_{scr} mice are phenotypically male but have one active and one inactive X chromosome. The mH2A1.2 staining of liver sections of XX_{scr} mice was identical to that of normal female mice (data not shown). Finally, we analysed a human skin cell line derived from a boy with Klinefelter's syndrome¹⁷. The sex chromosome complement of this cell line is XXXXY (one active and three inactive X chromosomes). Of these nuclei, 63% had three MCBs each (Fig. 4a, b). We also observed preferential mH2A1.2 staining of three chromosomes in metaphase spreads prepared from these cells

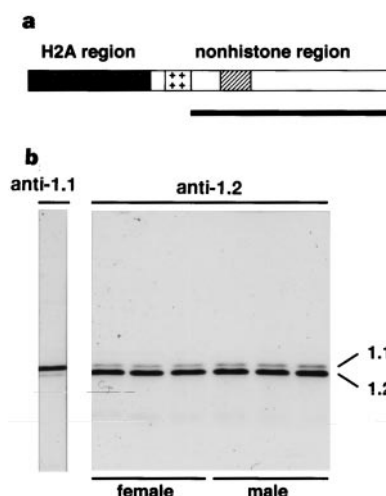


Figure 1 Specificity of mH2A antibodies. **a**, A diagram of mH2A1 subtypes. mH2A1.1 and 1.2 are identical apart from a segment generated by alternative splicing (cross-hatched). The fragment used to generate the antibodies is indicated by the solid bar. The H2A region is shaded; the segment rich in basic amino acids is indicated by plus signs. **b**, Western blot analysis of mouse liver nuclear extracts using antibodies raised against mH2A1.1 (anti-1.1) and mH2A1.2 (anti-1.2). The mH2A1.2 blot shows extracts from six littermates. Gel loading was normalized against core-histone content¹².

(Fig. 4c). MCBs were not detected in normal male primary skin fibroblasts (data not shown).

We conclude that mH2A1 protein is preferentially concentrated in the inactive X chromosome. Our antibodies against mH2A1.2 do not detect MCBs in every female cell. This may in part reflect technical limitations of our immunofluorescence detection. However, in some cells, antibodies against mH2A2.2, a subtype encoded by a second mH2A gene, detect MCBs that are unreactive to antibodies against mH2A1.2 (our unpublished results). Also, rearrangements of chromatin structure during replication may interfere with the detection of MCBs in dividing cells. Some inactive X

chromosomes may not be associated with any mH2A subtype. This would be consistent with the observation that other known mechanisms involved in X inactivation, including *Xist* expression, are not always required for maintenance of the inactive state^{18–20}.

Our results demonstrate the preferential association of a protein with the inactive X chromosome. This preference does not reflect a general association of mH2A1 protein with heterochromatin. In mouse liver nuclei, most of the heterochromatic domains detected with fluorescent DNA dyes did not stain heavily with our antibodies against mH2A (Fig. 2b). The specificity of mH2A1 histones for the inactive X chromosome and the evolutionary conservation of this association in humans, mice and dogs indicate that mH2A may function in the inactivation process. We suggest that the high concentration of mH2A-containing nucleosomes (macronucleosomes) in the inactive X chromosome participates in the formation of a distinct chromatin structure (macrochromatin) involved in transcriptional silencing. It seems likely that macronucleosomes interact with other components of the inactive X chromosome, including possibly *Xist* RNA. Although the *Xist* gene is essential for initiation of the inactivation process^{21–24}, little is known about how it mediates gene silencing. *Xist* RNA does not seem to encode a protein^{9,10} but does coat the inactive X chromosome (ref. 25). *Xist* RNA could alter the chromatin structure of the inactive X through interaction with mH2A.

Although mH2A1 histones seem to be involved in X inactivation, their function is not confined to the inactive X chromosome as is indicated by the presence of mH2A1.2 staining at locations other than the inactive X in female and male nuclei. This includes the rare MCBs detected in male liver cells. These must involve chromatin domains other than the inactive X chromosome, and their occurrence appears to be cell-type-specific (our unpublished results). Chickens also have mH2A1.1 and 1.2 (J. R. P. and R. N. Fuji,

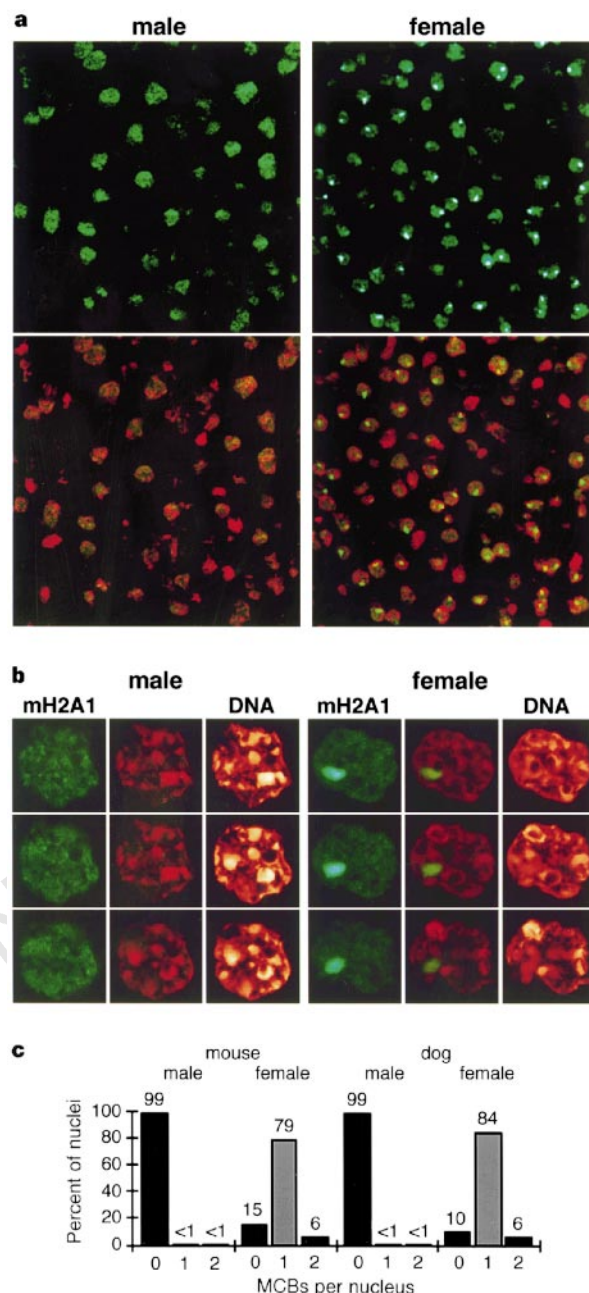


Figure 2 Sex-specific distribution of mH2A1 protein. **a**, Stacked confocal micrographs showing male and female mouse liver sections. Top panels, mH2A1.2 immunofluorescence (green); bottom panels, mH2A1.2 immunofluorescence (green) merged with DNA staining (red). **b**, Consecutive optical sections through a male and a female hepatocyte nucleus showing mH2A1.2 (green, left), DNA (orange, right) and the merged images (middle). **c**, Liver sections from mouse and dog were stained for mH2A1.2 and DNA. Intact nuclei were identified by confocal microscopy and MCBs were counted. Results are given as a percentage of total nuclei analysed (male mouse, 220; female mouse, 253; male dog, 280; female dog, 563).

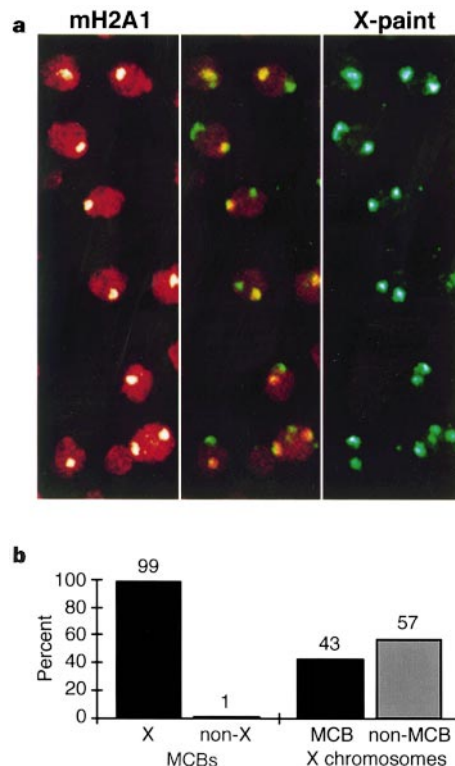


Figure 3 Colocalization of MCBs to X chromosomes. **a**, Stacked confocal micrographs of immunofluorescence/FISH results of female mouse hepatocyte nuclei showing mH2A1.2 staining (red), X-paint hybridization (green) and the merged image in between. **b**, Results of immunofluorescence/FISH given as a percentage of the total number of MCBs analysed (317) and of the total number of X chromosomes identified (730).

manuscript in preparation). As birds diverged from mammals before the development of X-chromosome inactivation, the association of mH2A1 histones with the inactive X could be an adaptation of a previously existing mechanism for modifying chromatin structure and function. One possibility is that macrochromatin is involved in gene silencing at other chromosomal locations. Our finding that the expression of mH2A1 subtypes differs in different cell types and changes during development¹³ indicates that mH2A may participate in the establishment and/or maintenance of patterns of gene expression during development. □

Methods

Antibody production. Antibodies against a bacterially produced glutathione S-transferase (GST) fusion protein containing the non-histone region of rat mH2A1.2 were raised in rabbits (Cocalico Biologicals), passed through a GST column, bound to an mH2A1.2 non-histone protein fragment column, eluted¹³ and passed through an mH2A2.2 column to eliminate crossreaction.

Animals and cells. Normal mice were of a variety of strains and ranged in age from 5 weeks to 5 months. Male and female dogs (*Canis familiaris*) were 4 months and 9 months old, respectively. XX_{Sxr} mice, X/0 mice, and their control siblings were from Jackson Labs. Human XXXXY skin fibroblasts (ATCC CCL-28, Dempsey) and normal male primary skin fibroblasts (ATCC CRL-2201) were grown on coverslips according to the supplier's instructions. Metaphase spreads were prepared by colcemid treatment, followed by trypsinization, hypotonic swelling, cytocentrifugation and fixation with paraformaldehyde.

Immunofluorescence. Frozen-tissue sections, 5–10 µm thick, were fixed

directly in 4% buffered paraformaldehyde for 5 min at room temperature, quenched in 100 mM Tris, pH 8, 150 mM NaCl, 0.5% Triton X-100 and blocked with 5% non-fat dry milk in TBSN (10 mM Tris, pH 7.4, 150 mM NaCl, 0.06% N-P40). Sections were incubated overnight at 4 °C with the primary antibody diluted in TBSN/milk (this incubation was for two nights in the case of cultured human fibroblasts because they have less mH2A1 protein), washed with TBSN/milk, incubated with the FITC-labelled mouse anti-rabbit IgG (Sigma), washed with TBSN/milk, counterstained with 10 mg ml⁻¹ propidium iodide in TBS, destained in TBS, mounted in Vectashield (vector Laboratories) and viewed with a Leica TCS laser scanning confocal microscope.

Immunofluorescence/FISH. Immunofluorescence was done as described, except that a Texas red donkey anti-rabbit IgG was used as secondary antibody and the DNA counterstain was omitted. Sections were refixed, then quenched in 100 mM Tris, pH 8, 150 mM NaCl, 0.5% Triton-X100 for 20 min, rinsed in TBS, fixed in 3:1 methanol:acetic acid for 20 min, washed in 2 × SSC for 30 min at 37 °C, washed again in 2 × SSC, incubated with 100 µg ml⁻¹ pepsin in 10 mM HCl for 8 min at 37 °C, washed in 2 × SSC, 0.1% SDS for 10 min, washed twice in 2 × SSC, dehydrated through a 70–85–100% ethanol series and air-dried. Hybridization was done using biotinylated mouse chromosome paint probes (Oncor) as recommended by the supplier. For hybrid detection, a layer of FITC-labelled streptavidin (Jackson ImmunoResearch) was followed by a layer of biotin-labelled anti-streptavidin (Accurate Chemical) and another layer of FITC-labelled streptavidin, as described¹⁵.

Received 18 February; accepted 27 March 1998.

1. Lyon, M. F. Gene action in the X-chromosome of the mouse (*Mus musculus* L.). *Nature* **190**, 372–373 (1961).
2. Heard, E., Clerc, P. & Avner, P. X-chromosome inactivation in mammals. *Annu. Rev. Genet.* **31**, 571–610 (1997).
3. Barr, M. L. & Bertram, E. G. A morphological distinction between neurones of the male and female, and the behaviour of the nucleolar satellite during accelerated nucleoprotein synthesis. *Nature* **163**, 676–677 (1949).
4. Taylor, J. H. Asynchronous duplication of chromosomes in cultured cells of Chinese hamster. *J. Biophys. Biochem. Cytol.* **7**, 455–464 (1960).
5. Jeppesen, P. & Turner, B. M. The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. *Cell* **74**, 281–289 (1993).
6. Brown, C. J. *et al.* A gene from the region of the human X inactivation centre is expressed exclusively from the inactive X-chromosome. *Nature* **349**, 38–44 (1991).
7. Borsani, G. *et al.* Characterization of a murine gene expressed from the inactive X-chromosome. *Nature* **351**, 325–329 (1991).
8. Brockdorff, N. *et al.* Conservation of position and exclusive expression of mouse Xist from the inactive X chromosome. *Nature* **351**, 329–331 (1991).
9. Brown, C. J. *et al.* The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized with the nucleus. *Cell* **71**, 527–542 (1992).
10. Brockdorff, N. *et al.* The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and is located in the nucleus. *Cell* **71**, 515–527 (1992).
11. Willard, H. F. & Salz, H. K. Remodelling chromatin with RNA. *Nature* **386**, 228–229 (1997).
12. Pehrson, J. R. & Fried, V. A. MacroH2A, a core histone containing a large nonhistone region. *Science* **257**, 1398–1400 (1992).
13. Pehrson, J. R., Costanzi, C. & Dharia, C. Developmental and tissue expression patterns of histone macroH2A1 subtypes. *J. Cell. Biochem.* **65**, 107–113 (1997).
14. Monestier, M., Losman, M. J., Novick, K. E. & Aris, J. P. Molecular analysis of mercury-induced antinuclear antibodies in H-2S mice. *J. Immunol.* **152**, 667–675 (1994).
15. Breneman, J. W. *et al.* The development of chromosome-specific composite DNA probes for the mouse and their application to chromosome painting. *Chromosoma* **102**, 591–598 (1993).
16. Evans, E. P., Burtenshaw, M. D. & Cattanach, B. M. Meiotic crossing-over between the X and Y chromosomes of male mice carrying the sex-reversing (Sxr) factor. *Nature* **300**, 433 (1982).
17. Hwa, T. C. & Lockhart, L. H. The beginning of the terminal stages of DNA synthesis of human cells with an XXXXY constitution. *Heredity* **52**, 320–324 (1965).
18. Migeon, B. R., Wolf, S. F., Axelman, J., Kaslow, D. C. & Schmidt, M. Incomplete X chromosome dosage compensation in chorionic villi of human placenta. *Proc. Natl Acad. Sci. USA* **82**, 3390–3394 (1985).
19. Kaslow, D. C. & Migeon, B. R. DNA methylation stabilizes X chromosome inactivation in eutherians but not in marsupials: evidence for multistep maintenance of mammalian X dosage compensation. *Proc. Natl Acad. Sci. USA* **84**, 6210–6214 (1987).
20. Brown, C. J. & Willard, H. F. The human X-inactivation centre is not required for maintenance of X chromosome inactivation. *Nature* **368**, 154–156 (1994).
21. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for Xist in X-chromosome inactivation. *Nature* **379**, 131–137 (1996).
22. Herzing, L. B. K., Romer, J. T., Horn, J. M. & Ashworth, A. Xist has properties of the X-chromosome inactivation centre. *Nature* **386**, 272–275 (1997).
23. Lee, J. T. & Jaenisch, R. Long-range cis effects of ectopic X-inactivation centres on a mouse autosome. *Nature* **386**, 275–279 (1997).
24. Marahrens, Y., Panning, B., Dausman, J., Strauss, W. & Jaenisch, R. Xist-deficient mice are defective in dosage compensation but not spermatogenesis. *Genes Dev.* **11**, 156–166 (1997).
25. Clemson, C. M., McNeil, J. A., Willard, H. F. & Lawrence, J. B. Xist RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J. Cell Biol.* **132**, 259–275 (1996).

Acknowledgements. We thank E. M. Eicher for the XX_{Sxr} mice; M. Monestier for the anti-fibrillar antibody; M. L. Atchison, E. L. F. Holzbaur, O. Jacenko, L. D. Peachey and J. D. Tucker for making their resources available; G. Gray-Board, E. A. Holleran, X. Lowe, M. Radic and J. D. Tucker for technical advice; and M. L. Atchison, N. G. Avadhani, J. Burch, L. H. Cohen and E. A. Holleran for comments on this manuscript. This work was supported by the NIH. Part of this work was done in the Biomedical Image Analysis Facility at the University of Pennsylvania, supported by the NIH. The dog material was supplied by A. Polesky and J. H. Wolfe, supported by the NIDDK.

Correspondence and requests for materials should be addressed to J.R.P. (e-mail: pehrson@vet.upenn.edu).

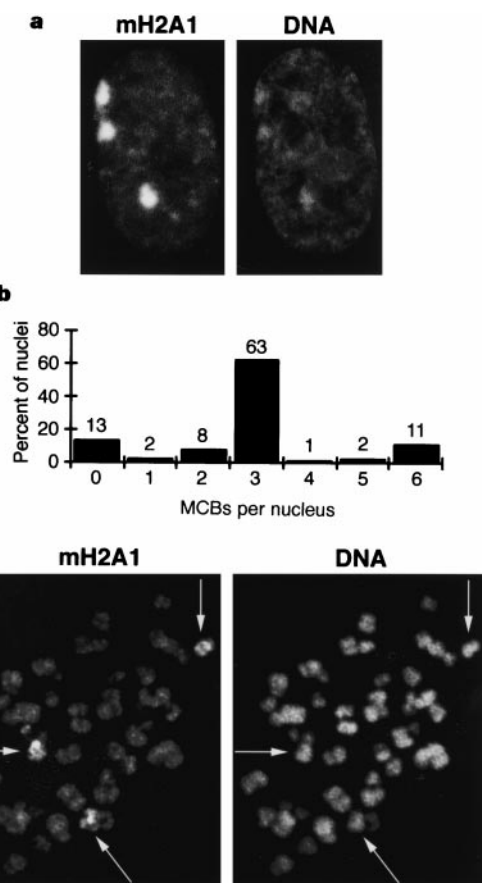


Figure 4 Correlation of MCBs with inactive X chromosomes. **a**, Stacked confocal micrographs showing a human XXXXY skin fibroblast nucleus stained for mH2A1.2 (left) and DNA (right). **b**, The number of MCBs per nucleus was quantified in human XXXXY skin fibroblasts. Results are presented as the percentage of total nuclei counted (193). **c**, Stacked confocal micrographs of a human XXXXY fibroblast metaphase spread stained for mH2A1.2 (left) and DNA (right). Arrows identify the three chromosomes that were preferentially stained with antibody against mH2A1.2.

Sequence cyberbiology

Tools for bioinformatics and computational biology featured here include servers and computation boxes, cross-platform software, new web-based interfaces to tools and databases, as well as proprietary and public data sets.

DeCypher I/ES series

From Time Logic

Bioinformatics accelerators designed for 'configurable computing'

A configurable computer is one in which tasks are accelerated by direct execution in hardware. However, the configurable computer's integrated circuits are able to rewire dynamically, as needed, to implement new or enhanced algorithms, thus avoiding the risk of rapid obsolescence associated with products employing fixed-logic chips (VLSI ASICs). The ES series is said to offer from 2,100 to over 210,000 acceleration beyond conventional workstations, performing computationally intensive algorithms, such as Smith-Waterman and FrameSearch. All models scale easily and are complete, ready-to-run network servers, eliminating the hassle and hidden expenses associated with 'add-on' boxes. The ES-series is stated to deliver 3.8 billion affine Smith-Waterman cells per second with quad SMP CPUs, RAM expansion to 2 GB and RAM cached, triple fast/wide Ultra-SCSI PCI disk system supporting over 100 gigabytes database storage. Additional reading on configurable computing: Villasenor, J. & Mangione-Smith, W. H. Configurable Computing. *Sci. Am.* 66-71 (June 1997).

Reader Enquiry No. 100

Octane and Origin systems

From Silicon Graphics

Multiprocessor systems for such algorithms as BLAST, FASTA and Smith-Waterman

The high performance and scalability of the Octane, Origin200, Origin2000 and CrayOrigin2000 systems are the result of Silicon Graphics' fast R10000 processor and supporting I/O subsystem, combined with highly tuned codes that maintain good locality of data. BLAST, FASTA and Smith-Waterman searches take advantage of coarse grain parallelism that is used in these systems, and are generally not limited by cache size. In the case of FASTA, the scalability actually increases with greater selectivity (lower k-tuple values) and/or larger databases, both of which increase the parallel workload and resultant scalability to multiple processors. According to the manufacturer, the Origin systems are well-suited to meeting the computational demands of biological sequence analyses using BLAST, FASTA and dynamic programming (Smith-Waterman).

Reader Enquiry No. 101

BioXL/G and BioXL/P

From Compugen

A new generation of hardware accelerators for genetic bioinformatic applications

These new systems use the latest field-programmable gate array (FPGA) and digital signal processing (DSP) technologies to deliver a combination of speed and accuracy. Product development was based on feedback from over 40 Compugen customers worldwide. The new accelerators are scaleable in configuration and meet the needs of demanding applications. The versatile BioXL/G accelerates most commonly used algorithms for database searching.

Reader Enquiry No. 102

GeneXus

From GeneXus

A PC-based client/server development environment for creating AS/400 applications

GeneXus is a PC-based product used to design and generate native AS/400 applications, using a knowledge-based application development approach. This knowledge base stores the components of the application and contains a comprehensive set of rules, which allow it to create a normalized database from a set of user definitions (screens, reports, rules, and the like). These are extracted during the application definition process. An application can be built and tested in an interactive mode on the PC. GeneXus implements advanced concepts from artificial intelligence and database research to create highly normalized data models automatically from simple, end-user object-oriented definitions. It controls objects and their relationships and generates efficient code on the targeted hardware. Moreover, its knowledge base export and import facility permits team integration of application components. GeneXus contains an integrated 4GL, rooted in mainframe technologies (CA-IDEAL, Natural/2), which facilitates the creation of complex programs and migration from other platforms.

Reader Enquiry No. 103

GeneMatcher

From Paracel

A network computation server designed for high-throughput bioinformatics environments

According to the manufacturer, the GeneMatcher system features a massive, parallel-pipeline architecture consisting of more than 6,000 custom ASIC processors,



Stacking up against the rest: Paracel's GeneMatcher server is built for speed.

which accelerate the execution of highly sensitive dynamic programming algorithms. The wide variety of supported algorithms allows researchers to rapidly compare similarities and differences between DNA and protein sequence data. GeneMatcher performs Smith-Waterman, frame search, profile search and hidden Markov-model searches at speeds that are stated to be thousands of times faster than traditional computing platforms, and $\times 40$ faster than Paracel's previous product, the FDF 3. The easy-to-use BioView Toolkit integrates GeneMatcher with any bioinformatics environment, says Paracel, and provides both end-user and application interfaces, including Java, HTML and UNIX command-line, with C and C++ callable functions.

Reader Enquiry No. 104

Bioinformatics analysis

bioScout

From Lion

An integrated software system for automated, high-throughput bioinformatics analyses

This system integrates a large number of bioinformatics tools to achieve comprehensive analysis of protein and nucleic acid sequences. The high degree of automation should make it possible to analyse entire genomes, as well as cDNA or EST collections, in shorter time frames. bioSCOUT also provides easy access to the results by analysing

and digesting the vast amount of output data, usually generated by bioinformatics programs. The entry port to this sequence information is a brief feature report, which summarizes the most important results, such as function and three-dimensional structure. This is accessed through a graphical user interface, which is based on web architecture. This interactive workbench guides the user to the appropriate bioinformatics tools for further analysis. The multi-user interactive environment allows users to share results. Daily updates of processed databases provide a continual stream of new information from all relevant, publicly available sources.

Reader Enquiry No. 105

SeqWeb

From Genetics Computer Group (GCG)
GCG, part of Oxford Molecular Group, offers a web-based interface to the 'Package'

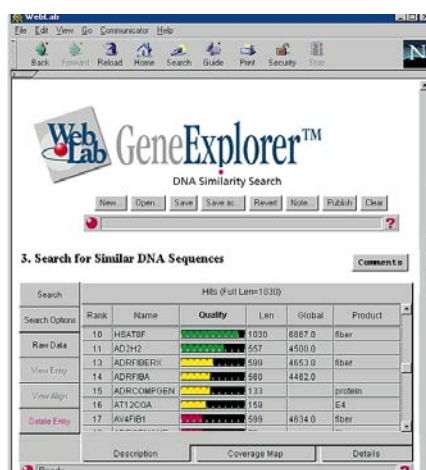
SeqWeb provides easy access to a core set of programs in the Wisconsin Package, which are now accessible through Netscape Navigator or Microsoft Internet Explorer, thus eliminating the need to memorize program commands and operating system specifics. Within SeqWeb, users can use sequence files from their desktop, access group project data on the server and search a site's standard and proprietary databases through a web browser. SeqWeb links to local databases, as well as those on the Internet, such as Medline, PDB, FlyBase, Mendelian Inheritance in Man (MIM) and the US Patent and Trademark Office patent database. There is also complete, context-sensitive help available online. In addition to SeqWeb, GCG also offers two other interfaces to the Wisconsin Package: SeqLab and the command line. Based on X Windows, SeqLab offers point-and-click access to Wisconsin Package programs and includes a powerful multiple sequence editor and annotation tool. The command line gives access to each program in a menu format and offers scriptable command lines for repetitive analyses. The Wisconsin Package provides over 130 DNA and protein analysis programs. One of the package's strengths is its access to a number of genetic databases and its powerful database searching capabilities.

Reader Enquiry No. 106

Gene Explorer 1.4

From Molecular Simulations (MSI)
New features help users understand the genetic basis of disease

The latest version of this computational environment for molecular biology and bioinformatics is now compatible with several major genomic database systems, including those from Incyte, Pangea Systems, PE Informatics, GCG, NCBI,



WebLab GeneExplorer stays in touch.

PDB and SRS. It automatically tells scientists when new database hits for previous experiments are found in database updates. In addition, the program helps researchers identify protein homologs with more sensitivity, using its protein threading technology. It does this by using secondary structure to augment conventional sequence similarity technologies. For more specific problems or questions, there is graphical access to parameter and algorithm controls, which makes it possible to fine-tune and optimize calculations to suit individual requirements.

Reader Enquiry No. 107

SPS products

From Southwest Parallel Software (SPS)
Based on three original Phil Green programs, the Oxford Molecular Group offers SPS Swat, Cross_Match and Phrap

These have been optimized for single and multi-CPU computer systems, and are available for SGI, DEC and Sun computers. Swat features are said to include excellent performance, scaling and support for multiple input files. The performance is stated to be 7.6 times faster when run on an 8-CPU computer, according to the manufacturer, and up to 19× faster on a 20-CPU computer. With Cross_Match, also available for SGI, DEC and Sun systems, the program runs up to 30 per cent faster than the original Cross_Match on a single-CPU system and up three times faster on a four-CPU system. This program has been optimized to work with all-to-all comparisons and a test sequence versus a database. Phrap now runs on these systems, as well, at speeds stated to be twice that of a single CPU computer system. (Scaling on multi-CPU computer systems is dependent on the data used.) For jobs that end in a multiple contig solution, Phrap will scale up to three times on a four-CPU computer system. SPS is currently working on developing a more optimized version of Phrap. Features should include

better memory management and parallelizing Phrap's merge routine.

Reader Enquiry No. 108

Database doubletake

PathoSeq

From Incyte

The release of version 2.0 of this microbial genome sequence database

The database contains genomic data for over a dozen microorganisms, along with bioanalysis tools that allow the exploration of microbial genomics and new drug targets for combating infectious disease. Incyte's high-throughput DNA sequencing program is said to produce fivefold coverage of a three megabase (Mb) genome in just three weeks, including at least 95 per cent coverage of all predicted genes. This allows the updating, including genomes nominated by subscribers. Six of the genomes are not available in the public domain: *Streptococcus pneumoniae*, *Staphylococcus aureus*, *Staphylococcus epidermidis*, *Enterococcus faecalis* and *Enterococcus faecium*. The June release of PathoSeq also includes Incyte's first fungal genome: *Candida albicans*.

Reader Enquiry No. 109

StackDB

From Pangea Systems

A human expressed gene dataset based on clustering ESTs into discrete gene records

This program provides an integrated dataset of the publicly available human ESTs and alignments. Developed at the South African National Bioinformatics Institute (SANBI), StackDB was designed to derive the longest current consensus representation of the expressed human genome. It should maximize the coverage of accurate, high-quality consensus sequences of human genes from available data to arrive at the best possible estimate of the makeup of the human gene set. This dataset features greater length and accuracy in its EST clustered consensi.

ADVERTISEMENT

Lysyl Endopeptidase
(Achromobacter Protease 1)
Specifically cleaves all Lys-X
bonds including Lys-Pro.
Highly Stable, Water Soluble

Wako

Wako Pure Chemical Industries, Ltd.
1-3-1, Nishikujo, Seto, Japan
Tel: (81) 562-27-7777
Fax: (81) 562-27-7777
E-mail: wako@wako-chem.com

READER ENQUIRY NO. 31